



# The Associations Between Consumption of Coffee and Soy Food With Health Outcomes.

## Citation

Ding, Ming. 2016. The Associations Between Consumption of Coffee and Soy Food With Health Outcomes.. Doctoral dissertation, Harvard T.H. Chan School of Public Health.

## Permanent link

<http://nrs.harvard.edu/urn-3:HUL.InstRepos:25757889>

## Terms of Use

This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at <http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA>

## Share Your Story

The Harvard community has made this article openly available.  
Please share how this access benefits you. [Submit a story](#).

[Accessibility](#)

**THE ASSOCIATIONS BETWEEN  
CONSUMPTION OF COFFEE AND SOY FOOD  
WITH  
HEALTH OUTCOMES**

Ming Ding

A Dissertation Submitted to the Faculty of  
The Harvard T.H. Chan School of Public Health  
in Partial Fulfillment of the Requirements  
for the Degree of Doctor of Science  
in the Departments of Nutrition and Epidemiology

Boston, Massachusetts.

March, 2016

Dissertation Advisor: Dr. Frank B. Hu

Ming Ding

## **THE ASSOCIATIONS BETWEEN CONSUMPTION OF COFFEE AND SOY FOOD WITH HEALTH OUTCOMES**

### **ABSTRACT**

Obesity has become a global epidemic, and obesity prevalence rose from 4.8% to 9.8% in men and from 7.9% to 13.8% in women between 1980 and 2008. Preventing obesity and related chronic diseases, especially type 2 diabetes (T2D) and cardiovascular disease (CVD), is of crucial public health significance. Identification of dietary factors that are beneficial to health is of high *priority*. This dissertation focused on two kinds of foods, coffee and soy food, and their associations with obesity-related health outcomes.

In Chapter 1, the dose-response relationship of long-term coffee consumption with CVD risk remained inconclusive. In the current study, I examined the association between coffee consumption and risk of CVD by meta-analyzing results from 36 prospective cohort studies with 1,279,804 study participants and 36,352 CVD cases. Our results showed that coffee consumption was non-linearly associated with risk of CVD: moderate coffee consumption was associated with lower risk of CVD, with the lowest CVD risk at 3 to 5 cups per day, and heavy coffee consumption was not associated with risk of CVD. However, whether the non-linear association was due to a true biological effect or confounding of smoking is not known. Therefore, in Chapter 2, with 208,501 participants and 31,956 deaths in three large cohort studies, I prospectively examined the associations of coffee consumption with total mortality and cause-specific mortality among the overall population as well as never smokers.

In Chapters 3 and 4, I examined the association of soy food with risk of type 2 diabetes. Two different approaches were used to assess soy food intake. First, soy food assessed by food frequency questionnaire (FFQ) was used as main exposure, and the association of soy food consumption with

risk of type 2 diabetes was examined prospectively in three Harvard cohorts. Second, urinary isoflavones excretion was used as main exposure, and the association of urinary isoflavones concentration with risk of type 2 diabetes was assessed using a nested case-control design.

## TABLE OF CONTENTS

<b>I. BODY of DISSERTATION</b>	<b>12</b>
Chapter 1. Long-term coffee consumption and risk of cardiovascular disease: a systematic review and a dose-response meta-analysis of prospective cohort studies	13
<i>Introduction</i>	15
<i>Methods</i>	16
<i>Results</i>	20
<i>Discussion</i>	24
<i>References</i>	29
Chapter 2. Association of coffee consumption with total and cause-specific mortality in three large prospective cohorts	69
<i>Introduction</i>	71
<i>Methods</i>	72
<i>Results</i>	76
<i>Discussion</i>	79
<i>References</i>	84
Chapter 3. Consumption of soy foods and isoflavones and risk of type 2 diabetes: a pooled analysis of three U.S. cohorts	131
<i>Introduction</i>	133
<i>Methods</i>	134
<i>Results</i>	139
<i>Discussion</i>	141
<i>References</i>	145

Chapter 4. Urinary isoflavones and risk of type 2 diabetes: a prospective investigation in U.S.

women 164

*Introduction* 166

*Study population and methods* 167

*Results* 171

*Discussion* 172

*References* 177

**II. CONCLUSION 193**

## LIST OF FIGURES

### **CHAPTER 1**

FIGURE 1. Study selection process of coffee consumption and risk of CVD.....	49
FIGURE 2. Forest plot of the associations between third, second, and highest level of coffee consumption and risk of CVD compared to the lowest level .....	50
FIGURE 3. Stratified analysis of the association between coffee consumption and risk of CVD	59
FIGURE 4. Dose response relationships of coffee consumption with risk of CVD.....	57
Supplemental Figure 1. Dose response relationship of coffee consumption with cardiovascular disease risk choosing only one outcome for correlated outcomes within the same study.....	62
Supplemental Figure 2. Dose response relationship of coffee consumption with cardiovascular disease risk from models adjusted for different confounders.....	63
Supplemental Figure 3. Egger's test for publication bias for the association between coffee consumption and risk of CVD.....	64

### **CHAPTER 2**

Figure 1. The association between coffee consumption and risk of mortality in the overall population and among never smokers pooled across the three cohorts.....	99
Figure 2. The association of a 1-cup per day increment in coffee consumption with risk of cause-specific mortality pooled across the three cohorts.....	102
Supplement Figure 1. Baseline coffee consumption and risk of mortality in the overall population and among never smokers in NHS.....	123
Supplemental Figure 2. Baseline coffee consumption and risk of mortality in the overall population and among never smokers in NHS .....	124
Supplemental Figure 3. Baseline coffee consumption and risk of mortality in the overall population and among never smokers in HPFS.....	125
Supplemental Figure 4. Cumulative coffee consumption and stopping updating when cancer and diabetes develop, and risk of mortality in the overall population and among never smokers by pooled across the three cohorts.....	126

Supplemental Figure 5. Cumulative coffee consumption and further stopping updating when hypertension, hypercholesterolemia develop, and risk of mortality in the overall population and among never smokers by pooled across the three cohorts.....	127
Supplemental Figure 6. Continue updating coffee consumption after diagnosis of chronic disease with 4-year lag and risk of mortality in the overall population and among never smokers pooled across the three cohorts.....	128
Supplemental Figure 7. Continue updating coffee consumption after diagnosis of chronic disease adjusting for hypercholesterolemia as a time varying covariates and risk of mortality in the overall population and among never smokers pooled across the three cohorts.....	129
Supplemental Figure 8. Baseline coffee consumption and risk of mortality in the overall population and among never smokers further excluding hypertension, hypercholesterolemia, and diabetes cases at baseline, pooled across the three cohorts.....	130

### **CHAPTER 3**

Supplemental Figure 1. The association between isoflavones consumption and risk of type 2 diabetes (T2D) in a dose response manner by pooling the three cohorts. ....	157
---	-----

### **CHAPTER 4**

Figure 1. The joint association of urinary isoflavones biomarkers and postmenopausal status and hormone use with risk of type 2 diabetes.....	186
---	-----



## LIST OF TABLES

### **CHAPTER 1**

Table 1. Basic characteristics of included studies.....	33
Supplemental table 1. the quality assessment of included studies using the Newcastle–Ottawa scale.....	58
Supplemental table 2. Egger’s test for the publication bias on coffee consumption and risk of type 2 diabetes.....	61

### **CHAPTER 2**

Table 1. Age-adjusted baseline characteristics of participants by frequency of total coffee consumption in NHS, NHS 2, and HPFS.....	88
Table 2. HRs (95% CI) for the association between consumption of total coffee, caffeinated coffee, and decaffeinated coffee and risk of mortality .....	90
Table 3. HRs (95% CI) for the association between consumption of total coffee, caffeinated coffee, and decaffeinated coffee and risk of mortality among never smokers.....	93
Table 4. Multivariate HRs (95% CI) for the association between consumption of total coffee and risk of cause-specific mortality among never smokers.....	96
Supplemental Table 1. Categories for causes of death.....	103
Supplemental Table 2. The disease composition of overall mortality.....	104
Supplemental Table 3. HRs (95% CI) for the association between consumption of total coffee and risk of cause-specific mortality.....	106
Supplemental Table 4. Stratified analysis for the association between coffee consumption and risk of total mortality among ever smoker.....	108
Supplemental Table 5. Stratified analysis for the association between coffee consumption and risk of total mortality.....	110
Supplemental Table 6. The tests of proportional hazard assumption in NHS, NHS 2, and HPFS.....	113
Supplemental Table 7. Sensitivity analyses for the association between coffee consumption and total mortality in the overall population, and pooled multivariate-adjusted hazard ratio was shown.....	115

Supplemental Table 8. Sensitivity analyses for the association between coffee consumption and total mortality among never smokers, and pooled multivariate-adjusted hazard ratio was shown.....119

Supplemental Table 9. The association between coffee consumption and risk of total mortality among never smokers by Cox model with inverse probability weighting.....122

### **CHAPTER 3**

Table 1. Baseline characteristics of participants by consumption of soy foods in the NHS, NHS II, and HPFS.....148

Table 2. Hazard ratios (HRs) for the associations between soy containing foods and risk of type 2 diabetes in the three cohorts.....150

Table 3. Associations between isoflavone consumption and risk of type 2 diabetes in the three cohorts.....153

Table 4. Hazard ratio (HR) for the association between subtypes of isoflavone consumption and risk of type 2 diabetes in the three cohorts.....155

Supplemental Table 1. Stratified analysis of the association between total soy food consumption (consumer vs. non-consumer) and risk of type 2 diabetes (T2D).....158

Supplemental Table S2. Stratified analysis of the association between consumption of isoflavones (dichotomized) and risk of type 2 diabetes (T2D).....159

Supplemental Table S3. Hazard ratio (HR) for the associations between residual isoflavones consumption after adjusting for coffee and risk of type 2 diabetes in the three cohorts.....160

Supplemental Table S4. Hazard ratios (HRs) for the associations between soy containing foods and risk of type 2 diabetes in the three cohorts using propensity score analysis.....162

Supplemental Table S5. Associations between isoflavone consumption and risk of type 2 diabetes in the three cohorts using propensity score analysis.....163

### **CHAPTER 4**

Table 1. The age-adjusted baseline characteristics according to diabetes cases and controls in the combined cohort.....180

Table 2. Odds ratio (95% CI) of type 2 diabetes by tertiles of urinary isoflavones (nmol/g creatinine) in the combined cohort.....182

Table 3. Stratified analysis of the association between urine isoflavones biomarkers and risk of type 2 diabetes by menopausal status and postmenopausal hormone use in the combined cohort.....	184
Supplemental Table 1. The age-adjusted baseline characteristics according to urinary total isoflavones in the controls of the combined cohort.....	188
Supplemental Table S2. Spearman correlation coefficient† between urine isoflavone biomarkers and soy food products assessed by FFQ among controls in the combined cohort.....	190
Supplemental Table S3. Stratified analysis of the association between urine isoflavones biomarkers and risk of type 2 diabetes by age, BMI, and AHEI.....	191

## ACKNOWLEDGEMENTS

---

I would like to express my sincere gratitude to my advisor, Dr. Frank B. Hu. Being mentored by Dr. Hu at Harvard is one of the most important opportunities in my life. I deeply appreciate his straightforward and insightful advice on how to become an excellent researcher. I thank him for leading me into the field of nutrition epidemiology and helping me lay a solid foundation for my future career.

Many colleagues helped me a lot with my research. I would like to express my appreciation to Dr. Qi Sun, Dr. Rob van Dam, Dr. Shilpa Bhupathiraju, Ambika Satija, Dr. Yanping Li, and Dr. Tao Huang, who kindly and collaboratively helped me make progress for my research projects. I also had productive collaborations with Dr. Lu Qi and Dr. Jorge E. Chavarro, and I thank their guidance in my research. I am also very grateful to Channing staff who are always ready to help.

I would like to thank my dissertation committee members, Dr. Bernard Rosner, Dr. Edward Giovannucci, and Dr. Qi Sun, for generosity in their time and for thoroughly reviewing my research with thoughtful and constructive comments.

My sincere thanks also go to my study group in the first two years and we learned a lot from each other: Sheng-Hsuan Lin, Wenyan Li, Mingyang Song, Xin Li, Tianyi Huang, and Kate Fitzgerald. I have shared joy and tears with my officemate Juan Wu for four years, and I cherish her encouragement during the tough times.

Finally, I would like to express my deepest gratitude to my family, friends, classmates, and former teachers for their support and encouragement.

## **I. BODY OF DISSERTATION**

## CHAPTER 1

---

# LONG-TERM COFFEE CONSUMPTION AND RISK OF CARDIOVASCULAR DISEASE: A SYSTEMATIC REVIEW AND A DOSE-RESPONSE META-ANALYSIS OF PROSPECTIVE COHORT STUDIES

---

Ming Ding MS, <sup>1</sup> Shilpa N Bhupathiraju PhD, <sup>1</sup> Ambika Satija BA, <sup>1</sup> Rob M van Dam PhD, <sup>1,2</sup>

Frank B Hu, MD, PhD <sup>1,3,4</sup>

<sup>1</sup>Department of Nutrition, Harvard School of Public Health, Boston, MA

<sup>2</sup>Saw Swee Hock School of Public Health and Department of Medicine, Yong Loo Lin School of Medicine, National University of Singapore and National University Health System, Singapore

<sup>3</sup>Department of Epidemiology, Harvard School of Public Health, Boston, MA

<sup>4</sup>Channing Division of Network Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA

## ABSTRACT

**Background** Considerable controversy exists regarding the association between coffee consumption and cardiovascular disease (CVD) risk. A meta-analysis was performed to assess the dose-response relationship of long-term coffee consumption with CVD risk.

**Methods and Results** Pubmed and EMBASE were searched for prospective cohort studies of the relationship between coffee consumption and CVD risk, which included coronary heart disease, stroke, heart failure, and CVD mortality. Thirty-six studies were included with 1,279,804 participants and 36,352 CVD cases. A non-linear relationship of coffee consumption with CVD risk was identified (P for heterogeneity = 0.09, P for trend < 0.001, P for non-linearity < 0.001). Compared with the lowest category of coffee consumption (median: 0 cups/d), the relative risk of CVD was 0.95 (95% CI, 0.87 to 1.03) for the highest (median: 5 cups/d) category, 0.85 (0.80 to 0.90) for the second highest (median: 3.5 cups/d), and 0.89 (0.84 to 0.94) for the third highest category (median: 1.5 cups/d). Looking at separate outcomes, coffee consumption was non-linearly associated with both CHD (P for heterogeneity = 0.001, P for trend < 0.001, P for non-linearity < 0.001) and stroke risks (P for heterogeneity = 0.07, P for trend < 0.001, P for non-linearity < 0.001) (P for trend differences > 0.05).

**Conclusions** A non-linear association between coffee consumption with CVD risk was observed in this meta-analysis. Moderate coffee consumption was inversely significantly associated with CVD risk, with the lowest CVD risk at 3 to 5 cups/d, and heavy coffee consumption was not associated with elevated CVD risk.

**Key Words:** coffee, cardiovascular disease, meta-analysis

## INTRODUCTION

Coffee is one of the most widely consumed beverages around the world; thus, investigating whether or not coffee consumption is associated with chronic disease risk has important public health implications. The relationship between coffee consumption and risk of coronary heart disease was first studied in the 1960s, given that the prevalence of coffee drinking and CHD were both high in western countries.<sup>1</sup> Short-term metabolic studies found that caffeine ingestion acutely induces cardiac arrhythmias, and increases plasma renin activity, catecholamine concentrations, and blood pressure.<sup>2,3</sup> In the 1980s, cross-sectional studies found a positive association between coffee consumption and serum total cholesterol concentrations, which might be related to the coffee brewing method (i.e. boiled or unfiltered coffee).<sup>4</sup> A later randomized trial showed that boiled coffee consumption increased the serum cholesterol.<sup>5</sup> From the 1980s to the 2000s, many case-control studies, which are prone to recall and selection bias, showed a positive association between coffee consumption and CHD risk.<sup>6-8</sup> In contrast, meta-analyses of prospective cohort studies tended to find no association, although results varied substantially across studies.<sup>9,10</sup>

Since 2000, the association between coffee consumption and other cardiovascular disease (CVD) outcomes such as stroke, heart failure, and total CVD mortality has also been more frequently studied.<sup>11-13</sup> Meta-analyses have been published to summarize the association between coffee and risk of CHD,<sup>14</sup> stroke,<sup>15</sup> and heart failure.<sup>16</sup> These meta-analyses did not support an association between coffee consumption and a higher CVD risk, but the shape of the association remains uncertain. Moreover, a number of additional studies have been published since the publication of these meta-analyses,<sup>11,13,17-19</sup> and one recent meta-analysis paper showed that heavy coffee consumption was not associated with risk of CVD mortality.<sup>20</sup> To



examine the dose response association of coffee consumption with cardiovascular disease risk, we conducted a systematic review and meta-analysis of coffee consumption and incidence of total CVD outcomes, including incidence of CHD, stroke, and heart failure, and CVD mortality.

## **Methods**

We followed the Meta-Analysis of Observational Studies in Epidemiology<sup>21</sup> protocol throughout the design, implementation, analysis, and reporting of our meta-analysis.

### **Search strategy and selection criteria**

We searched the PubMed and EMBASE databases for prospective studies that had evaluated the association between coffee consumption and risk of CVD between January 1966 and March 2013. The computer-based searches included the key words “coffee”, “cardiovascular disease”, “coronary heart disease”, “stroke”, “mortality”, “heart failure”, “myocardial infarction”, “ischemic heart disease”, “sudden cardiac arrest”, and “acute coronary syndrome”. Reference lists of retrieved articles were manually scanned for all relevant additional studies and review articles. We restricted the search to studies on humans that were written in English.

### **Study Selection**

Studies were included in this meta-analysis if they met the following criteria: 1) prospective cohort studies, including case-cohort studies and nested case-control studies with a prospective design; 2) the exposure was coffee consumption, including total coffee, caffeinated coffee, or decaffeinated coffee; 3) the outcome was risk of CVD, including incidence of CHD, stroke, and heart failure, and CVD mortality. Studies were excluded if 1) the study had a retrospective design; 2) the estimates were presented without standard errors or other information that allowed

calculation of standard errors; 3) the outcome was atrial fibrillation, atherosclerosis, hypertension, aortic stiffness, or venous thrombus; 4) no confounders were adjusted for.

### **Data extraction and quality assessment**

One author (M. D.) assessed study eligibility and extracted the data, the other author (A. S.) independently double-checked the available data. The following data were extracted from each study: first author's name, year of publication, geographical location, follow-up time, sex, age, number of CVD events, number of participants/person-years of follow up, categories of coffee consumption, mean/median coffee consumption in each category, CVD assessment method, covariates adjusted for in the multivariable analysis, and relative risks and the associated measure of variance for all categories of coffee consumption. For cohorts with published data on several CVD outcomes, we chose incidence instead of mortality or heart failure results. For studies with data on both CHD and stroke as the outcome, we included both in the meta-analysis. The correlation of CHD and stroke was accounted for in the main analysis (see below). In a sensitivity analysis, we analyzed one of the two outcomes. The Newcastle-Ottawa quality assessment scale (NOS) <sup>22</sup> was used to evaluate the quality of the included studies. M. D. and S. B. developed the evaluation criteria (supplemental table 1). The score ranges from 0 to 9 points with a higher score indicating higher study quality.

To perform a dose-response meta-analysis, we assigned the median coffee consumption in each category of consumption to the corresponding relative risk for each study. We used means for this purpose if medians were not reported. If neither the mean nor the median consumption per category was reported, the midpoint of the upper and lower boundaries in each category was used to estimate median consumption. If the upper boundary for the highest category was not provided, the assigned median value was 25% higher than the lower boundary

of that category. If the lower boundary for the lowest category was not provided, the assigned median value was half of the upper boundary of that category.

### **Data Synthesis and Analysis**

To analyze the trend of coffee consumption and risk of CVD, we used both semi-parametric and parametric methods. For the semi-parametric method, four coffee consumption groups were generated, namely lowest, third highest, second highest, and highest. For each study that was included, the lowest and the highest coffee consumption categories corresponded to the lowest and highest groups, respectively. For studies with four exposure categories, the second and third categories corresponded to the second and third highest groups, respectively. For studies with three exposure categories, the middle category corresponded to either the second or the third highest group in the meta-analysis, depending on the similarity of the median coffee consumption to either the second or the third highest group of the meta-analysis. If the study had more than four exposure categories, two consumption groups, other than the lowest and highest, were chosen based on their similarity of the amount of coffee consumption in that category to the second and third highest groups of the meta-analysis. For each group, we computed correlation coefficients ( $\rho$ ) between CHD and stroke outcomes in the same cohort. We imputed  $\rho = 1$  initially to obtain the most conservative effect estimates. A random-effects model was used first and was changed to a fixed-effects model if no between study heterogeneity was found for the random-effects model ( $\text{tau-squared} < 1$ ).<sup>23</sup> Sensitivity analysis was conducted by imputing different  $\rho$  ( $0 < \rho \leq 1$ ) to evaluate the robustness of the effect estimates. We used the STATA command ROBUMETA to obtain the effect estimates.

For the parametric method, a dose-response meta-analysis was performed.<sup>24</sup> The number of cases and participants in each coffee consumption category was extracted to estimate the

covariance of the relative risk in each study. Together with the observed adjusted variance of the relative risk, we estimated the variance/covariance matrix of the data. The weight of each study was calculated as the inverse of the variance/covariance matrix. We used generalized least squares models (GLST) with the maximum likelihood method to estimate the coefficients for each study. We fit a fixed-effects generalized linear model first, and changed to a random-effects generalized linear model if the p value for the goodness of fit/heterogeneity of the previous model was  $< 0.05$ . Additionally, we tested for potential non-linearity in the association between coffee consumption and CVD risk using a fixed/random-effects restricted cubic spline model with 3 knots. In sensitivity analysis, we used two-stage fixed/random-effects dose response models to combine studies that reported results for categorized coffee consumption and studies with reported results for continuous coffee consumption. Specifically, the RR of CVD per unit increase of coffee consumption for each study was first estimated separately by GLST, and then the RRs from all of the studies were pooled together by a fixed/random-effects model. We used the STATA command GLST for model fitting, and the command LINCOME to obtain effect estimates for the fitted model.

We performed stratified analyses by baseline hypertension or MI of the study population, smoking status, publication year, NOS study quality score, dietary assessment method, evaluation of stroke or CHD as the outcome, country, sex, and type of coffee (caffeinated coffee or decaffeinated coffee). The interaction between categorized coffee consumption and the stratifying variable with the risk of CVD was tested by a likelihood ratio test comparing the models derived using GLST method with and without the interaction terms. We assessed the potential for publication bias using Egger's regression symmetry test.<sup>25</sup> All analyses were conducted using STATA Version 11.2 (STATA Corp, College Station, Texas).

## Results

### Characteristics of studies

Our initial search identified 2587 potentially relevant citations. After screening titles and abstracts, we identified 53 studies for further evaluation. Of the 53 initially included studies, we excluded 14 studies due to duplicate publication, one study with point estimate without standard error, and one nested case control study with a retrospective design. Thirty-six studies remained in the meta-analysis (Figure 1). The included studies comprised approximately 1,283,685 study participants and 47,779 CVD cases, including 28,347 CHD cases, 12,030 stroke cases and 7,402 other CVD cases. Characteristics of these 36 studies are shown in Table 1. One study had a nested case-control study design, one had a case-cohort study design, and the rest of the studies were cohort studies. Duration of follow-up for incident CVD ranged from 6 to 44 years, with a median follow-up of 10 years. Twenty-one studies were conducted in Europe, 12 in the US, and 3 in Japan. Three studies assessed coffee consumption repeatedly during the course of the follow-up, and the rest of the studies assessed coffee consumption at baseline. Thirteen studies assessed coffee consumption without using a specific dietary assessment method, and the rest of the studies assessed coffee consumption by diet recalls, diet records or food frequency questionnaires (FFQ). One study modeled coffee consumption as a continuous variable, and the remaining studies modeled coffee consumption categorically. Nine studies assessed the association of caffeinated coffee consumption with CVD risk, and four studies assessed the association of decaffeinated coffee consumption with CVD risk. The outcome in 17 studies was risk of stroke, while the outcome in 22 studies was risk of CHD. The scores of the NOS quality assessment ranged from 3 to 8, and 31 studies had scores of 5 or higher. The corresponding

results of each criteria of the NOS quality assessment for our meta-analysis are shown in Supplemental Table 1. The study modeling coffee as a continuous exposure was excluded in the following analysis due to the difficulty of combining the risk estimate with those of other studies and was only included in the sensitivity analysis.<sup>26</sup> All of the remaining 35 studies were included in the main analysis, and 29 studies were included in the dose-response analysis between coffee consumption and risk of CVD.

### **Coffee consumption and risk of CVD**

The relative risks for CVD with different coffee consumption categories relative to the lowest category are shown in Figure 2. Of the 35 studies, 6 cohorts presented the outcome of stroke and CHD simultaneously. Compared with the lowest category of coffee consumption (median and mean: 0 cups/d), the pooled RR for incident CVD was 0.89 (95% CI, 0.84 to 0.94) for the third highest (median: 1.5 cups/d; mean: 1.48 cups/d), 0.85 (95% CI, 0.80 to 0.90) for the second highest (median: 3.5 cups/d; mean: 3 cups/d) and 0.95 (95% CI, 0.87 to 1.03) for the highest (median: 5 cups/d; mean: 5.5 cups/d) category of coffee consumption (Figure 2). Low between-study variances of CVD risk were found for each category of coffee consumption (tau-squared = 0.00 for the random-effects models), and the imputed correlation coefficient between the risks of stroke and CHD within the same cohort ( $0 < \rho \leq 1$ ) did not have an effect on the relative risk of CVD for each category of coffee consumption.

### **Stratified analyses**

Stratified analyses were conducted according to baseline hypertension or MI of the study population, smoking status, publication year, NOS study quality score, dietary assessment method (24-h diet recall/diet record/FFQ versus other methods), stroke versus CHD as the

outcome, country, sex, and type of coffee (caffeinated coffee or decaffeinated coffee). No interactions between categorized coffee consumption and stratification variables in relation to CVD risk were observed (all  $P$  for interactions  $>0.05$ ) (Figure 3). Only 4 studies provided the stratified results by age.<sup>27-30</sup> The summarized results showed that, comparing the highest with the lowest intakes, the RR of CVD was 0.96 (95% CI, 0.65 to 1.42) for age  $< 65$  years, and the RR was 0.91 (95% CI, 0.59 to 1.40) for age  $\geq 65$  years.

For the risk of CHD, compared with the lowest category of coffee consumption, the RRs of CHD were 0.89 (95% CI, 0.85 to 0.94;  $P$  for heterogeneity = 0.83;  $I^2 = 0.0\%$ ) for the third highest category, 0.90 (95% CI, 0.84 to 0.97;  $P$  for heterogeneity = 0.02;  $I^2 = 40.3\%$ ) for the second highest category, and 0.93 (95% CI, 0.84 to 1.02;  $P$  for heterogeneity  $< 0.001$ ;  $I^2 = 52.8\%$ ) for the highest category of coffee consumption. The corresponding RRs of stroke were 0.89 (95% CI, 0.84 to 0.94;  $P$  for heterogeneity = 0.58;  $I^2 = 0.0\%$ ) for the third category, 0.80 (95% CI, 0.75 to 0.86;  $P$  for heterogeneity = 0.37;  $I^2 = 6.5\%$ ) for the second category, and 0.95 (95% CI, 0.84 to 1.07;  $P$  for heterogeneity = 0.001;  $I^2 = 54.5\%$ ) for the highest category.

### **Dose-response analysis of coffee consumption with risk of CVD**

In our dose-response analysis, we observed a non-linear association between coffee consumption and risk of CVD ( $P$  for non-linearity  $< 0.001$ ) with a significant trend ( $P$  for trend  $< 0.001$ ) and limited heterogeneity in study results ( $P$  for heterogeneity = 0.09) (Figure 4a). Compared to those with no coffee consumption, the RR estimated directly from the cubic spline model was 0.95 (95% CI, 0.93 to 0.97) for 1 cup/d, 0.92 (95% CI, 0.88 to 0.95) for 2 cups/d, 0.89 (95% CI, 0.85 to 0.93) for 3 cups/d, 0.88 (95% CI, 0.83 to 0.93) for 4 cups/d, 0.89 (95% CI, 0.83 to 0.95) for 5 cups/d, 0.91 (95% CI, 0.84 to 0.99) for 6 cups/d, and 0.93 (95% CI, 0.85 to 1.03) for 7 cups/d.

Non-linear (p values for non-linearity  $<0.001$ ) associations between coffee consumption and disease risk with significant trends (p values for trend  $<0.001$ ) were found for both CHD and stroke (Figures 4b and 4c). There was stronger evidence for heterogeneity in study results for the association of coffee consumption with CHD risk (P heterogeneity=0.001) than for the association with stroke risk (P heterogeneity=0.07).

We further explored the reason for the heterogeneity between coffee consumption and CHD risk by stratifying the studies by publication year ( $\leq 2000$  or  $> 2000$ ). We found that in studies published in year 2000 or earlier, coffee consumption was not significantly associated with CHD risk (n=13, P for heterogeneity = 0.20), whereas in later studies, coffee consumption was non-linearly associated with CHD risk (n= 18, P for heterogeneity = 0.08). We didn't perform a similar analysis for stroke because very few studies on stroke were published prior to 2000.

### **Sensitivity analysis**

We tested the robustness of our results in sensitivity analyses. Because the RRs of stroke and CHD from the same cohort were correlated and a total of 6 studies included both CHD and stroke results, we conducted a sensitivity analysis by including only one outcome at a time. Our results remained largely unchanged and non-linear curves were found with including either CHD or stroke as the outcome (Supplemental Figure 1a and 1b).

One study with coffee consumption modeled as a continuous variable was excluded from the main analysis;<sup>26</sup> we added the RR from this study to the dose-response analysis by a two stage method and the results did not substantially change.

To test whether the association between coffee consumption and risk of CVD was different for unadjusted and multivariable adjusted models, we performed a dose-response meta-analysis of the only age-adjusted data including 34 comparisons (Supplemental Figure 2). Multivariate



adjustment strengthened the inverse association between moderate consumption and CVD risk, most likely due to adjustment for smoking.

### **Publication bias**

The Egger test did not suggest publication bias for associations for any category of coffee consumption and risk of CVD (Supplemental Figure 3 and Table 2).

## **DISCUSSION**

The findings from this systematic review and meta-analysis, based on approximately 1,283,685 study participants and 47,779 CVD cases, including about 28,347 CHD cases, 12,030 stroke cases and 7,402 other CVD cases, demonstrate a non-linear association between coffee consumption and risk of CVD. Moderate coffee consumption (3-5 cups/day) was associated with lower CVD risk, and heavy coffee consumption ( $\geq 6$  cups/day) was neither associated with a higher nor a lower risk of CVD.

In contrast to our results, a previous meta-analysis summarizing 21 prospective cohort studies<sup>31</sup> found no association between moderate coffee consumption and CHD risk in the overall population. One possible reason is that the previous meta-analysis included 7 studies without adjustment for confounders, which might have biased the relative risks upwards because of confounding by factors such as smoking.

A recent cohort study by Liu et al<sup>32</sup> found that 4 cups per day of coffee consumption was associated with increased mortality, but the association was only significant for participants under 55 years old. The results from this study contradict those from this meta-analysis and the majority of studies in the literature. Possible reasons for this discrepancy include a relatively

small size, lack of updated dietary assessment, and subgroup analysis. In our meta-analysis, stratified analysis by age revealed no significant differences in the association across age groups.

The debate about the relation between coffee consumption and CVD risk mainly stemmed from inconsistent results according to different study designs. Case-control studies, which are prone to recall bias and selection bias, tended to show a positive association, whereas cohort studies generally showed a null association.<sup>10</sup> Still, findings from prospective cohort studies on coffee consumption and CVD risk have remained inconsistent. Differences among studies in sample sizes, the characteristics of the study populations, the assessment methods for coffee consumption, and statistical adjustments may have contributed to divergent results. Since the true association between coffee consumption and CVD risk is likely to be modest and nonlinear, the differences in coffee assessments and covariate adjustments may result in changes the magnitude and even the direction of the associations and thus lead to different conclusions.

The U-shaped association between coffee consumption and CVD risk observed in this meta-analysis need to be considered from both methodological and biological points of view. First, individuals with hypertension or other conditions related to CVD risk might have changed their coffee consumption before baseline. Thus, baseline disease, especially hypertension, as a confounder could result in reverse causation. However, we observed no significant difference in the association between coffee consumption and CVD risk between cohorts with hypertensive and MI patients and the general population cohorts. Second, smoking is likely to be an important confounder for the association between coffee consumption and CVD risk, and could bias the relative risks upwards. Heavy coffee consumption was associated with higher risk of CVD in age-adjusted analyses, but this is likely due to confounding by smoking. After adjustment for

smoking and other covariates, heavy coffee consumption was not significantly associated with CVD and the inverse association between moderate consumption and CVD became stronger.

The non-linear U-shaped association between coffee consumption and CVD risk might also be true based on plausible biological mechanisms. Coffee is a complex chemical mixture with hundreds of compounds including the phenolic compound chlorogenic acid, caffeine, minerals such as potassium and magnesium, niacin and its precursor trigonelline, and lignans. Coffee consumption has been associated with higher insulin sensitivity, a lower risk of type 2 diabetes, and lower concentrations of inflammatory markers such as C-reactive protein and E-selectin.<sup>33, 34</sup> However, short-term metabolic studies have shown that caffeine can acutely increase blood pressure by antagonizing the adenosine A1 and A2A receptor,<sup>35-37</sup> and could also acutely adversely affect arterial stiffness and endothelium dependent vasodilation.<sup>38, 39</sup> Long-term heavy coffee consumption has been associated with a slightly elevated risk of hypertension,<sup>40</sup> and a higher level of plasma homocysteine.<sup>41, 42</sup> In addition, cafestol in unfiltered coffee increases serum total cholesterol concentrations.<sup>43</sup> The non-linear U-shaped association between coffee consumption and risk of CVD might be due to a combination of beneficial and detrimental effects: for moderate coffee consumption, beneficial effects may be greater than adverse effects; whereas for heavy consumption, detrimental effect may counterbalance beneficial effects. Results from case crossover studies suggest that coffee consumption transiently increases risk of nonfatal myocardial infarction, ischemic stroke onset, and sudden cardiac death.<sup>44-46</sup> However, we could not differentiate acute effects from long-term effects of habitual coffee consumption in this study.

No significant association between decaffeinated coffee consumption with CVD risk was observed in this meta-analysis. There were several potential explanations. First, the consumption

of decaffeinated coffee was much lower than caffeinated coffee, diminishing the power to detect any association. Second, the null association might be due to a reverse causation problem in that individuals with hypertension or other CVD-related conditions might switch from regular coffee to decaffeinated coffee. This reverse causation may mitigate an inverse association between decaffeinated coffee consumption and CVD risk.

We did not observe a significant association between coffee consumption and CHD risk for earlier publications (2000 or earlier). There are two potential reasons for this finding. First, coffee brewing methods have changed over time and nowadays the filter method has become more popular, effectively replacing unfiltered forms of coffee such as boiled coffee that was more widely consumed by participants in earlier studies. It has been shown that drinking boiled coffee increases serum cholesterol, an important risk factor for CVD <sup>5</sup>. Second, in earlier studies, the sample size was typically small; the measurement of baseline characteristics was typically crude; statistical control of confounders such as diet was inadequate; and the average NOS study quality score was lower. Our stratified analysis showed that coffee consumption was not associated with CVD risk in subgroups with a lower NOS score.

A study by Cornelis et al. <sup>47</sup> showed that CYP1A2 genotype was an effect modifier between coffee consumption and risk of myocardial infarction: coffee consumption was related to higher risk of myocardial infarction for the slow caffeine metabolizer, and was not related to myocardial infarction for the fast caffeine metabolizer. However, this analysis was based on a case-control study conducted in Costa Rico and the results have not been replicated in prospective cohort studies yet.

Recently, a genome-wide association study (GWAS) found a highly significant association between a variant on CYP1A2 and coffee intake <sup>48</sup>. However, this variant explains

only a very small population variance. Since the vast majority of our participants were Caucasians, the allele frequency was expected to be consistent across various cohorts. Ideally, the meta-analyses should be done according to different genotypes of CYP1A2. However, none of the included cohorts assessed the genotypes and thus we were unable to conduct such a stratified analysis.

Our meta-analysis has several strengths. First, our meta-analysis included 35 cohort studies and 1,283,685 participants, which provided sufficient power to detect modest associations. Second, because of the prospective design of all included studies, differential misclassification of coffee consumption due to recall bias was minimized and the likelihood of selection bias is reduced. Third, we used both semi-parametric and parametric methods, and both analyses indicated a U-shaped relationship between coffee consumption and CVD risk. Finally, we conducted stratified analyses according to disease endpoints, geographic locations of the studies, type of coffee, and baseline characteristics of the study populations. The subgroup results are highly consistent and robust.

Our study also has several limitations. Given the observational nature of the studies, the possibility of residual confounding cannot be excluded. However, since higher coffee consumption was generally associated with a less healthy lifestyle such as a higher prevalence of cigarette smoking, less physical activity, and a less healthy diet, the observed association between moderate coffee consumption and a lower CVD risk is unlikely to be explained by these confounders. In addition, residual confounding by smoking may have biased the association for heavy coffee consumption upward, which may explain our finding that adjustment for smoking and other covariates actually strengthened the inverse association. Nonetheless, because of the observational nature of the included studies, a causal relationship cannot be established with

these data alone. In addition, coffee brewing methods were not assessed in the included studies. However, given coffee consumption habits in the studied populations most consumed coffee is likely to have been filtered coffee. As a result, our results may not apply to unfiltered coffee (e.g. French press, Scandinavian boiled, or Turkish/Greek coffee).

In conclusion, our meta-analysis suggests a non-linear relationship between coffee consumption and CVD risk. Moderate coffee consumption was associated with lower CVD risk, with the lowest CVD risk at 3 to 5 cups/d of coffee consumption, and heavy coffee consumption was not associated with CVD risk. This non-linear association with coffee consumption was observed for both the risk of CHD and stroke.

**Funding sources** — NIH grant HL60712.

**Conflict of Interest Disclosures** — Dr. van Dam received a research grant from the Nestec Company. No potential conflicts of interest relevant to this article were reported.

### **Reference**

1. Paul O, Lepper MH, Phelan WH, Dupertuis GW, Macmillan A, Mc KH, Park H. A longitudinal study of coronary heart disease. *Circulation*. 1963;28:20-31
2. Robertson D, Frolich JC, Carr RK, Watson JT, Hollifield JW, Shand DG, Oates JA. Effects of caffeine on plasma renin activity, catecholamines and blood pressure. *N Engl J Med*. 1978;298:181-186
3. Dobmeyer DJ, Stine RA, Leier CV, Greenberg R, Schaal SF. The arrhythmogenic effects of caffeine in human beings. *N Engl J Med*. 1983;308:814-816
4. Thelle DS, Heyden S, Fodor JG. Coffee and cholesterol in epidemiological and experimental studies. *Atherosclerosis*. 1987;67:97-103
5. Bak AA, Grobbee DE. The effect on serum cholesterol levels of coffee brewed by filtering or boiling. *N Engl J Med*. 1989;321:1432-1437
6. Coffee drinking and acute myocardial infarction. Report from the boston collaborative drug surveillance program. *Lancet*. 1972;2:1278-1281
7. Jick H, Miettinen OS, Neff RK, Shapiro S, Heinonen OP, Slone D. Coffee and myocardial infarction. *N Engl J Med*. 1973;289:63-67

8. Hennekens CH, Drolette ME, Jesse MJ, Davies JE, Hutchison GB. Coffee drinking and death due to coronary heart disease. *N Engl J Med.* 1976;294:633-636
9. Kawachi I, Colditz GA, Stone CB. Does coffee drinking increase the risk of coronary heart disease? Results from a meta-analysis. *Br Heart J.* 1994;72:269-275
10. Greenland S. A meta-analysis of coffee, myocardial infarction, and coronary death. *Epidemiology.* 1993;4:366-374
11. Freedman ND, Park Y, Abnet CC, Hollenbeck AR, Sinha R. Association of coffee drinking with total and cause-specific mortality. *N Engl J Med.* 2012;366:1891-1904
12. Kokubo Y, Iso H, Saito I, Yamagishi K, Yatsuya H, Ishihara J, Inoue M, Tsugane S. The impact of green tea and coffee consumption on the reduced risk of stroke incidence in japanese population: The japan public health center-based study cohort. *Stroke.* 2013;44:1369-1374
13. Floegel A, Pischon T, Bergmann MM, Teucher B, Kaaks R, Boeing H. Coffee consumption and risk of chronic disease in the european prospective investigation into cancer and nutrition (epic)-germany study. *Am J Clin Nutr.* 2012;95:901-908
14. Sofi F, Conti AA, Gori AM, Eliana Luisi ML, Casini A, Abbate R, Gensini GF. Coffee consumption and risk of coronary heart disease: A meta-analysis. *Nutr Metab Cardiovasc Dis.* 2007;17:209-223
15. Larsson SC, Orsini N. Coffee consumption and risk of stroke: A dose-response meta-analysis of prospective studies. *Am J Epidemiol.* 2011;174:993-1001
16. Mostofsky E, Rice MS, Levitan EB, Mittleman MA. Habitual coffee consumption and risk of heart failure: A dose-response meta-analysis. *Circ Heart Fail.* 2012;5:401-405
17. Sugiyama K, Kuriyama S, Akhter M, Kakizaki M, Nakaya N, Ohmori-Matsuda K, Shimazu T, Nagai M, Sugawara Y, Hozawa A, Fukao A, Tsuji I. Coffee consumption and mortality due to all causes, cardiovascular disease, and cancer in japanese women. *J Nutr.* 2010;140:1007-1013
18. de Koning Gans JM, Uiterwaal CS, van der Schouw YT, Boer JM, Grobbee DE, Verschuren WM, Beulens JW. Tea and coffee consumption and cardiovascular morbidity and mortality. *Arterioscler Thromb Vasc Biol.* 2010;30:1665-1671
19. Larsson SC, Virtamo J, Wolk A. Coffee consumption and risk of stroke in women. *Stroke.* 2011;42:908-912
20. Malerba S, Turati F, Galeone C, Pelucchi C, Verga F, La Vecchia C, Tavani A. A meta-analysis of prospective studies of coffee consumption and mortality for all causes, cancers and cardiovascular diseases. *Eur J Epidemiol.* 2013
21. Stroup DF, Berlin JA, Morton SC, Olkin I, Williamson GD, Rennie D, Moher D, Becker BJ, Sipe TA, Thacker SB. Meta-analysis of observational studies in epidemiology: A proposal for reporting. Meta-analysis of observational studies in epidemiology (moose) group. *JAMA.* 2000;283:2008-2012
22. Wells GA SB, O'Connell D, Peterson J, Welch V, Losos M, Tugwell P. The newcastle-ottawa scale (nos) for assessing the quality of nonrandomized studies in meta-analysis. [www.ohri.ca/programs/clinical\\_epidemiology/oxford.asp](http://www.ohri.ca/programs/clinical_epidemiology/oxford.asp). 2011
23. Hedges LV TE, Johnson MC. Robust variance estimation in meta-regression with dependent effect size estimates. *Research Synthesis Methods.* 2010;1:39-65
24. Greenland S, Longnecker MP. Methods for trend estimation from summarized dose-response data, with applications to meta-analysis. *Am J Epidemiol.* 1992;135:1301-1309
25. Egger M, Davey Smith G, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. *BMJ.* 1997;315:629-634
26. Wilhelmsen L, Tibblin G, Elmfeldt D, Wedel H, Werko L. Coffee consumption and coronary heart disease in middle-aged swedish men. *Acta Med Scand.* 1977;201:547-552

27. Klatsky AL, Friedman GD, Armstrong MA. Coffee use prior to myocardial infarction restudied: Heavier intake may increase the risk. *Am J Epidemiol.* 1990;132:479-488
28. Lindsted KD, Kuzma JW, Anderson JL. Coffee consumption and cause-specific mortality. Association with age at death and compression of mortality. *J Clin Epidemiol.* 1992;45:733-742
29. Bidel S, Hu G, Qiao Q, Jousilahti P, Antikainen R, Tuomilehto J. Coffee consumption and risk of total and cardiovascular mortality among patients with type 2 diabetes. *Diabetologia.* 2006;49:2618-2626
30. Greenberg JA, Dunbar CC, Schnoll R, Kokolis R, Kokolis S, Kassotis J. Caffeinated beverage intake and the risk of heart disease mortality in the elderly: A prospective analysis. *Am J Clin Nutr.* 2007;85:392-398
31. Wu JN, Ho SC, Zhou C, Ling WH, Chen WQ, Wang CL, Chen YM. Coffee consumption and risk of coronary heart diseases: A meta-analysis of 21 prospective cohort studies. *Int J Cardiol.* 2009;137:216-225
32. Liu J, Sui X, Lavie CJ, Hebert JR, Earnest CP, Zhang J, Blair SN. Association of coffee consumption with all-cause and cardiovascular disease mortality. *Mayo Clin Proc.* 2013
33. van Dam RM, Hu FB. Coffee consumption and risk of type 2 diabetes: A systematic review. *JAMA.* 2005;294:97-104
34. Williams CJ, Fargnoli JL, Hwang JJ, van Dam RM, Blackburn GL, Hu FB, Mantzoros CS. Coffee consumption is associated with higher plasma adiponectin concentrations in women with or without type 2 diabetes: A prospective cohort study. *Diabetes Care.* 2008;31:504-507
35. Nurminen ML, Niittynen L, Korpela R, Vapaatalo H. Coffee, caffeine and blood pressure: A critical review. *Eur J Clin Nutr.* 1999;53:831-839
36. Mesas AE, Leon-Munoz LM, Rodriguez-Artalejo F, Lopez-Garcia E. The effect of coffee on blood pressure and cardiovascular disease in hypertensive individuals: A systematic review and meta-analysis. *Am J Clin Nutr.* 2011;94:1113-1126
37. Fredholm BB, Battig K, Holmen J, Nehlig A, Zvartau EE. Actions of caffeine in the brain with special reference to factors that contribute to its widespread use. *Pharmacol Rev.* 1999;51:83-133
38. Karatzis E, Papaioannou TG, Aznaouridis K, Karatzi K, Stamatelopoulos K, Zampelas A, Papamichael C, Lekakis J, Mavrikakis M. Acute effects of caffeine on blood pressure and wave reflections in healthy subjects: Should we consider monitoring central blood pressure? *Int J Cardiol.* 2005;98:425-430
39. Papamichael CM, Aznaouridis KA, Karatzis EN, Karatzi KN, Stamatelopoulos KS, Vamvakou G, Lekakis JP, Mavrikakis ME. Effect of coffee on endothelial function in healthy subjects: The role of caffeine. *Clin Sci (Lond).* 2005;109:55-60
40. Zhang Z, Hu G, Caballero B, Appel L, Chen L. Habitual coffee consumption and risk of hypertension: A systematic review and meta-analysis of prospective observational studies. *Am J Clin Nutr.* 2011;93:1212-1219
41. Verhoef P, Pasma WJ, Van Vliet T, Urgert R, Katan MB. Contribution of caffeine to the homocysteine-raising effect of coffee: A randomized controlled trial in humans. *Am J Clin Nutr.* 2002;76:1244-1248
42. Olthof MR, Hollman PC, Zock PL, Katan MB. Consumption of high doses of chlorogenic acid, present in coffee, or of black tea increases plasma total homocysteine concentrations in humans. *Am J Clin Nutr.* 2001;73:532-538
43. Urgert R, Katan MB. The cholesterol-raising factor from coffee beans. *Annu Rev Nutr.* 1997;17:305-324
44. Baylin A, Hernandez-Diaz S, Kabagambe EK, Siles X, Campos H. Transient exposure to coffee as a trigger of a first nonfatal myocardial infarction. *Epidemiology.* 2006;17:506-511



45. Selb Semerl J, Selb K. Coffee and alcohol consumption as triggering factors for sudden cardiac death: Case-crossover study. *Croat Med J*. 2004;45:775-780
46. Mostofsky E, Schlaug G, Mukamal KJ, Rosamond WD, Mittleman MA. Coffee and acute ischemic stroke onset: The stroke onset study. *Neurology*. 2010;75:1583-1588
47. Cornelis MC, El-Sohemy A, Kabagambe EK, Campos H. Coffee, cyp1a2 genotype, and risk of myocardial infarction. *JAMA*. 2006;295:1135-1141
48. Cornelis MC, Monda KL, Yu K, Paynter N, Azzato EM, Bennett SN, Berndt SI, Boerwinkle E, Chanock S, Chatterjee N, Couper D, Curhan G, Heiss G, Hu FB, Hunter DJ, Jacobs K, Jensen MK, Kraft P, Landi MT, Nettleton JA, Purdue MP, Rajaraman P, Rimm EB, Rose LM, Rothman N, Silverman D, Stolzenberg-Solomon R, Subar A, Yeager M, Chasman DI, van Dam RM, Caporaso NE. Genome-wide meta-analysis identifies regions on 7p21 (ahr) and 15q24 (cyp1a2) as determinants of habitual caffeine consumption. *PLoS Genet*. 2011;7:e1002033

Table1. Basic characteristics of included studies

Author/ Year/ Country/ Special annotation	Sex	Follow -up years	Age at start of follow- up (y)	No. of cases/Tot al No. of participa nts	Exposure(cup/d) Relative risk (95% CI)	Outcome	Exposure/ou tcome assessment	Confounders adjusted for
Wilhelmsen et al 1977 Europe	Men	12	50	60/834	Per cup increase of coffee consumption: 1.11 (0.83-1.51)	CHD	Not specific diet questionnaire (baseline)/ Hospital record	Smoking, cholesterol, SBP, dyspnea, registration by temperance board
Legrady et al 1987 US	men	19	40-56	220 CHD,57 stroke/ 1910	Stroke mortality	stroke mortality	Not specific diet questionnaire (baseline)/ Death certificates	age, diastolic blood pressure, serum cholesterol, and smoking status
					0-1cup/d 1.00 (1.00-1.00)			
					>1cup/d 1.64 (0.80-3.38)			
Martin et al 1988 US Hypertensive population	both	4	30-69	336/ 10,064	Stroke mortality	CVD mortality	Not specific diet questionnaire (baseline)/ Death certificates	age, sex, race, type of care, marital status, month of interview, body weight, initial diastolic blood pressure, fasting plasma blood glucose and serum cholesterol,
					0 cup/d 1.00 (1.00-1.00)			
					0.1mg-2 cups/d 0.73 (0.37- 1.46)			
					2-4 cups/d 0.61 (0.26-1.44)			
					>4 cups/d 1.30 (0.56-3.04)			
					CHD mortality			

					0 cup/d 1.00 (1.00-1.00)			initial end organ damage, and location of the study center
					0.1-2 cups/d 0.93 (0.66-1.3)			
					2-4 cups/d 0.81 (0.53-1.23)			
					>4 cups/d 0.80 (0.46-1.39)			
Grobbee et al 1990 US	men	2	40-75	411/45,589	0 cup/d 1.00 (1.00-1.00)	CVD	FFQ (baseline)/ Confirmed cases	age, quintiles of Quetelet's index, smoking habits, history of diabetes, alcohol use, parental history of myocardial infarction, specific health profession, energy intake, cholesterol, and saturated, monounsaturated, and polyunsaturated fat
					0-1 cup/d 0.70 (0.51-0.97)			
					2-3 cups/d 1.00 (0.79-1.26)			
					≥4 cups/d 0.90 (0.67-1.22)			
Klatsky et al 1990 US Nested case-control study	both	8 (median: 5)	From <50 to >60	1914/1,01,774	MI	CHD	Not specific diet questionnaire (baseline)/ Hospitalization for coronary disease	age, race, cigarette smoking, alcohol intake, education, baseline disease, and tea use.
					0 cup/d 1.00 (1.00-1.00)			
					<1cup/d 0.78 (0.56-1.07)			
					1-3cups/d 1.16 (0.93-1.45)			
					≥4 cups/d 1.42 (1.11-1.81)			
					Other coronary cases			
					0 cup/d 1.00 (1.00-1.00)			
					<1cup/d 0.90 (0.72-1.11)			
					1-3cups/d 0.89 (0.76-1.04)			
					≥4 cups/d 1.03 (0.85-1.24)			
Tverdal et al 1990 Europe	both	6.4	35-54	184/38564	no sugar in coffee	CHD mortality	Not specific diet questionnaire (baseline)/ Confirmed cases	age, high density lipoprotein, total cholesterol, systolic blood pressure, no of cigarettes/day
					<1 cup/d 1.00 (1.00-1.00)			
					≥9 cups/d 4.10 (1.30-13.20)			
					sugar in coffee			
					<1 cup/d 1.00 (1.00-1.00)			
					≥9 cups/d 1.60 (0.60-4.30)			
Rosengren et	men	7.1	51-59	399/	0 cup/d 1.00 (1.00-1.00)	CHD	Not specific	age, systolic blood

al 1991 Europe				6765	≥9 cups/d 1.40 (0.80-2.40)		diet questionnaire (baseline)/ National registries	pressure, body mass index, diabetes, registration for alcohol abuse, family history of myocardial infarction, mental stress, physical activity, and occupational class, smoking
Lindsted et al 1992 Europe	men	15	≥30	NA/ 9484	<1 cup/d 1.00 (1.00-1.00)	CVD mortality	FFQ (baseline)/ Confirmed cases	Body mass index, stroke, heart disease, hypertension, race, exercise, sleep, marital status, education, smoking history, dietary pattern
					1-2 cups/d 1.38 (1.18-1.62)			
					≥3 cups/d 1.44 (1.18-1.76)			
Klag et al 1994 US	men	32	26	111/ 1040	0 cup/d 1.00 (1.00-1.00)	CHD	Not specific diet questionnaire (baseline)/ National registries	age at graduation, baseline serum cholesterol, calendar time, time-dependent hypertension status, number of cigarettes, diabetes, and body mass index
					1-2 cups/d 1.70 (0.78-3.68)			
					3-4 cups/d 3.02 (1.37-6.65)			
					≥5 cups/d 2.94 (1.27-6.81)			
Gyntelberg et al 1995 Europe	men	6	53-74	184/ 2975	1-4 cups/d 1.00 (1.00-1.00)	CHD	Not specific diet questionnaire (baseline)/ Confirmed cases	age, alcohol, blood pressure, serum selenium level, social class, and triglycerides
					5-8 cups/d 1.00 (0.70-1.40)			
					≥9 cups/d 0.60 (0.30-1.00)			
Hart et al 1997 Europe	men	21	35-64	625/ 5766	0 cup/d 1.00 (1.00-1.00)	CHD mortality	Not specific diet questionnaire (average)/	age, diastolic blood pressure, cholesterol, smoking, social class, age leaving full time education,
					0.5-1 cup/d 1.20 (0.87-1.64)			
					1.5-2 cup/d 1.17 (0.83-1.65)			
					2.5-4 cup/d 1.16 (0.81-1.66)			

					>4.5 cup/d 1.49 (0.89-2.47)		National registries	body mass index, angina, and ECG ischaemia
Hakim et al 1998 US Hypertensive population	men	25	55-68	76/499	0 cup/d 1.00 (1.00-1.00) ≥6 cups/d 2.1 (1.2-3.7)	Stroke	24h diet recall (baseline)/ Confirmed cases	age, systolic blood pressure, total cholesterol, triglycerides, diabetes, alcohol use, and the physical activity index as measured at the time of study enrollment
Woodward et al 1999 Europe	both	7.7	40-59	567/ 11000	Men 0 cup/d 1.00 (1.00-1.00) 1-2 cups/d 0.68 (0.42-1.10) 3-4 cups/d 0.39 (0.21-0.73) ≥5 cups/d 0.68 (0.37-1.24) women 0 cup/d 1.00 (1.00-1.00) 1-2 cups/d 0.54 (0.22-1.34) 3-4 cups/d 0.56 (0.20-1.56) ≥5 cups/d 0.55 (0.18-1.66)	CHD	Food consumption table (baseline)/ Confirmed cases	age, housing tenure, activity at work, activity in leisure, cigarette smoking status, body mass index, Bortner score, cotinine, systolic blood pressure, fibrinogen, total cholesterol, HDL-cholesterol, triglycerides, alcohol, vitamin C, and tea.
Kleemola et al 2000 Europe	both	10	30-59	1645/ 20179	men with nonfatal MI <1 cup/d 1.09 (0.78-1.54) 1-3 cups/d 1.00 (1.00-1.00) 4-7 cups/d 0.95 (0.79-1.15) >7 cups/d 0.79 (0.64-0.98) women with nonfatal MI <1 cup/d 1.72 (1.01-2.92) 1-3 cups/d 1.00 (1.00-1.00) 4-7 cups/d 0.84 (0.62-1.13) >7 cups/d 0.93 (0.63-1.36)	CHD, CHD mortality	Not specific diet questionnaire (baseline)/ National registries	age, smoking status, serum cholesterol level, blood pressure, and history of MI

					men with CHD mortality <1 cup/d 1.88 (1.20-2.95) 1-3 cups/d 1.00 (1.00-1.00) 4-7 cups/d 1.23 (0.93-1.62) >7 cups/d 1.22 (0.90-1.65) women with CHD mortality <1 cup/d 0.00 (0.00-0.00) 1-3 cups/d 1.00 (1.00-1.00) 4-7 cups/d 0.67 (0.41-1.07) >7 cups/d 0.57 (0.28-1.16)			
Jazbec et al 2003 Europe	both	10	35-59	435/ 3364	men 0 cup/d 1.00 (1.00-1.00) <1 cups/d 0.70 (0.49-1.00) 1-2 cups/d 0.82 (0.58-1.16) >2 cups/d 0.72 (0.44-1.18) women 0 cup/d 1.00 (1.00-1.00) <1 cup/d 0.88 (0.56-1.39) 1-2 cups/d 0.67 (0.43-1.05) >2 cups/d 0.62 (0.30-1.28)	CVD mortality	Not specific diet questionnaire (baseline)/ Confirmed cases	age, number of cigarettes consumed per day, diastolic blood pressure, ulcer, feeling of well- being, region
Happonen et al 2004 Europe	Men	14	42-60	269/1971	None 0.84 (0.41-1.72) Light 1.22 (0.90-1.64) Moderate 1.00 (1.00-1.00) Heavy 1.43 (1.06-1.94)	CHD mortality	Diet record (baseline)/ National registries	Age, packyears of smoking, ischemia in exercise test, diabetes, income, and serum insulin concentration. Physical activity; family history of CHD; intake of alcohol, tea, saturated fat, total energy, and total water; serum

								glucose and plasma vitamin C concentration
Lopez- Garcia et al 2006 US	both	20	men 53 women 46	4427/ 128493	Women	CHD	FFQ (baseline)/ National registries	age, smoking status, serum cholesterol level, blood pressure, and history of MI
					<0.033 cup/d 1.00 (1.00-1.00)			
					0.033-0.57 cup/d 0.97 (0.83-1.14)			
					0.57-1 cup/d 1.02 (0.90-1.17)			
					2-3 cups/d 0.84 (0.74-0.97)			
					4-5 cups/d 0.99 (0.83-1.17)			
					≥6 cups/d 0.87 (0.68-1.11)			
					Men			
					<0.033 cup/d 1.00 (1.00-1.00)			
					0.033-0.57 cup/d 1.04 (0.91-1.17)			
					0.57-1 cup/d 1.02 (0.90-1.15)			
					2-3 cups/d 0.97 (0.86-1.11)			
					4-5 cups/d 1.07 (0.88-1.31)			
					≥6 cups/d 0.72 (0.49-1.07)			
Andersen et al 2006 US	wo men	15	55-69	1411/ 27312	0 cup/d 1.00 (1.00-1.00)	CVD mortality	FFQ (baseline)/ National registries	age, smoking, and intake of alcohol, BMI, waist-hip ratio, education, physical activity, use of estrogens,
					<1 cup/d 0.85 (0.68-1.06)			
					1-3 cups/d 0.76 (0.64-0.91)			
					4-5 cups/d 0.81 (0.66-0.99)			

					≥6 cups/d 0.87 (0.69-1.09)			use of multivitamin supplements, energy intake, and intakes of whole and refined grain, red meat, fish and seafood, and total fruit and vegetables
Bidel et al 2006 Europe Type 2 diabetic population	both	20.8	25-74	909/3837	0-2 cups/d 1.00 (1.00-1.00)	CVD mortality	Not specific diet questionnaire (baseline)/ National registries	age, sex, study year, BMI, systolic blood pressure, total cholesterol, education, alcohol and tea consumption, and smoking status
					3-4 cups/d 0.79 (0.64-0.97)			
					5-6 cups/d 0.70 (0.57-0.86)			
					≥7 cups/d 0.71 (0.56-0.90)			
Greenberg et al 2007 US	both	8.8	32-86	426/6594	<65y	CVD mortality	FFQ (baseline)/ Confirmed cases	age, smoking, BMI, sex, race, physical activity, alcohol consumption, per capita income, educational level, and American-style diet
					<0.5 cup/d 1.00 (1.00-1.00)			
					0.5-2 cups/d 0.95 (0.38-2.35)			
					2-4 cups/d 0.79 (0.34-1.85)			
					≥4 cups/d 0.86 (0.38-1.06)			
					≥65y			
					<0.5 cup/d 1.00 (1.00-1.00)			
					0.5-2 cups/d 0.72 (0.52-0.99)			
					2-4 cups/d 0.69 (0.52-0.92)			
					>4 cups/d 0.53 (0.38-0.75)			
Silletta et al 2007 Europe	both	3.5	52-63	1167/11231	0 cup/d 1.00 (1.00-1.00)	CVD	FFQ (average)/ Confirmed	Age, gender, smoking, BMI, dietary habits, cardiovascular risk
					<2 cups/d 1.02 (0.87-1.20)			
					2-4 cups/d 0.91 (0.75-1.09)			



Myocardial infarction population					>4 cups/d 0.88 (0.64-1.2)		cases	factors, history of MI before the index MI, time from the index MI to enrollment, post-MI complications, and pharmacological therapies, with inclusion of the allocation treatments
Greenberg et al 2008 US	both	10.1	65-97	523/1354	0 cup/d 1.00 (1.00-1.00) ≥1 cup/d 1.00 (0.84- 1.20)	CVD events	FFQ (baseline)/ Confirmed cases	age, gender, smoking, body mass index, alcohol consumption, physical activity, marital status, BP, history of CVD, and antihypertensive medication use
Happonen et al 2008 Europe	both	14.5	70-94	344/817	0 cups/d 0.80 (0.47-1.35) 1-2 cups/d 1.00 (1.00-1.00) 3-4 cups/d 0.96 (0.72-1.27) 5-6 cups/d 0.89 (0.64-1.24) ≥7 cups/d 0.84 (0.48-1.47)	CVD mortality	Not specific diet questionnaire (baseline)/ National registries	sex, current age, calendar period, marital status, educational level, previous occupational group, current smoking, BMI, history of myocardial infarction, presence of diabetes mellitus, cognitive impairment, physical disability, and self-rated health
Larsson et al 2008 Europe	men	13.6	50-69	2702/26556	<2 cup/d 1.00 (1.00-1.00) 2-3 cups/d 0.91 (0.79-1.06) 4-5 cups/d 0.88 (0.77-1.02) 6-7 cups/d 0.77 (0.66-0.90) ≥8 cups/d 0.77 (0.66-0.90)	Stroke	FFQ (baseline)/ National registries	age, supplementation group, No. of cigarettes smoked daily, body mass index, systolic and diastolic blood pressure, serum total cholesterol, serum HDL cholesterol,

								histories of diabetes and coronary heart disease, leisure-time physical activity, alcohol intake, and tea consumption
Mukamal et al 2009 Europe Myocardial infarction population	both	6.9-9.9	45-70	331(MI), 135(stroke)/ 1369	MI	CHD, stroke	FFQ (baseline)/ National registries	age, sex, diabetes, smoking, obesity, physical inactivity, alcohol consumption, tea consumption, education, and intake of boiled coffee
					0-1 cup/d 1.00 (1.00-1.00)			
					1-3 cups/d 0.97 (0.65-1.45)			
					3-5 cups/d 0.75 (0.50-1.13)			
					5-7 cups/d 0.94 (0.61-1.44)			
					≥7 cups/d 0.84 (0.51-1.40)			
					Stroke			
					0-1 cup/d 1.00 (1.00-1.00)			
					1-3 cups/d 1.08 (0.57-2.02)			
					3-5 cups/d 0.94 (0.49-1.78)			
					5-7 cups/d 1.17 (0.59-2.29)			
					≥7 cups/d 0.74 (0.31-1.75)			
Sugiyama et al 2010 Japan	both	10.3	40-64	426/ 37742	men CVD mortality	CVD mortality	FFQ (baseline)/ Mortality certificates at the public health center	age in years, sex, past history of hypertension and diabetes, education level, BMI, walking time, cigarette smoking, consumption of alcohol, green tea, oolong tea, black tea, intake of rice, miso soup, total meat, total dairy products, total fish, total vegetables, total fruits, and energy
					0 cup/d 1.00 (1.00-1.00)			
					0-1 cup/d 1.09 (0.79-1.51)			
					1-2 cups/d 0.85 (0.56-1.23)			
					≥3 cups/d 0.88 (0.56-1.39)			
					women CVD mortality			
					0 cup/d 1.00 (1.00-1.00)			
					0-1 cup/d 0.56 (0.36-0.86)			
					1-2 cups/d 0.48 (0.29-0.80)			
					≥3 cups/d 0.45 (0.20-1.03)			
Ahmed et al	men	9	45-79	784/	≤1 cup/d 1.00 (1.00-1.00)	Heart	FFQ	age, body mass index, total

2009 Europe				37315	2 cups/d 0.87 (0.69-1.11)	failure	(baseline)/ Confirmed cases	activity score, smoking, history of high cholesterol, family history of MI before age 60, education level, marital status, aspirin use, alcohol, tea, energy-adjusted fat intake, and energy-adjusted daily sodium intake
					3 cups/d 0.89 (0.70-1.14)			
					4 cups/d 0.89 (0.69-1.15)			
					≥5 cups/d 0.89 (0.69-1.15)			
Lopez- Garcia et al 2009 US	wo men	24	56	2280/ 83076	<0.03 cup/d 1.00 (1.00-1.00)	Stroke	FFQ (average)/ Confirmed cases	age, smoking status, body mass index, physical activity, alcohol intake, menopausal status and use of hormone replacement therapy, aspirin use; total caloric intake; quintiles of calcium, potassium, sodium, and folate intake; glycemic load; whole grain intake; and tertiles of fruits, vegetables, and fish consumption, high blood pressure, hypercholesterolemia, and type 2 diabetes mellitus
					0.03-0.57 cup/d 0.96 (0.82-1.13)			
					0.57-1 cup/d 0.88 (0.77-1.02)			
					2-3 cups/d 0.84 (0.72-0.98)			
					≥4 cups/d 0.85 (0.69-1.06)			
Leurs et al	both	10	55–69	1789(IH)	men with MI mortality	CHD	FFQ	age, current smoking,

2010 Europe Case-cohort study				D deaths), 708(strok e deaths)/ 120852	0-2 cups/d 1.00 (1.00-1.00)	mortality, stroke mortality	(baseline)/ National registries	number of cigarettes smoked, years of active smoking and total energy intake
					2-4 cups/d 0.91 (0.71-1.16)			
					3-6 cups/d 1.02 (0.79-1.31)			
					>6 cups/d 1.17 (0.86-1.59)			
					women with MI mortality			
					0-2 cups/d 1.00 (1.00-1.00)			
					2-4 cups/d 0.75 (0.58-0.97)			
					3-6 cups/d 0.62 (0.46-0.84)			
					>6 cups/d 0.71 (0.45-1.12)			
					men with stroke mortality			
					0-2 cups/d 1.00 (1.00-1.00)			
					2-4 cups/d 0.84 (0.60-1.18)			
					3-6 cups/d 0.72 (0.50-1.04)			
					>6 cups/d 1.15 (0.74-1.77)			
					women with stroke mortality			
					0-2 cups/d 1.00 (1.00-1.00)			
					2-4 cups/d 0.79 (0.57-1.09)			
					3-6 cups/d 0.70 (0.48-1.02)			
					>6 cups/d 1.10 (0.63-1.90)			
Gans et al 2010 Europe	both	13	20-69	1387(CH D cases), 563(strok e cases), 123(CHD deaths), 70(stroke deaths)/ 37514	CHD morbidity	CHD, CHD mortality, Stroke, stroke mortality	FFQ (baseline)/ National registries	sex; age; educational level; physical activity; smoking status; waist circumference; menopausal status; alcohol, tea; total energy; and saturated fat, fiber, and vitamin C level
					<1 cup/d 1.00 (1.00-1.00)			
					1-2 cups/d 0.85 (0.70-1.04)			
					2-3 cups/d 0.79 (0.65-0.96)			
					3-4 cups/d 0.82 (0.68-0.98)			
					4-6 cups/d 0.86 (0.73-1.02)			
					>6 cups/d 0.91 (0.74-1.11)			
					stroke morbidity			
					<1 cup/d 1.00 (1.00-1.00)			
					1-2 cups/d 1.08 (0.79-1.47)			

					2-3 cups/d 1.15 (0.85-1.57)			
					3-4 cups/d 1.10 (0.82-1.46)			
					4-6 cups/d 1.11 (0.84-1.46)			
					>6 cups/d 1.22 (0.88-1.70)			
					CHD mortality			
					<1 cup/d 1.00 (1.00-1.00)			
					1-3 cups/d 1.06 (0.61-1.84)			
					3-6 cups/d 0.64 (0.37-1.11)			
					>6 cups/d 0.73 (0.37-1.42)			
					Stroke mortality			
					<1 cup/d 1.00 (1.00-1.00)			
					1-3 cups/d 0.86 (0.39-1.87)			
					3-6 cups/d 1.20 (0.59-2.47)			
					>6 cups/d 1.34 (0.49-3.64)			
Larsson et al 2011 Europe	wo men	10.4	49-83	1680/ 34670	<1 cup/d 1.00 (1.00-1.00) 1-2 cups/d 0.78 (0.66-0.91) 3-4 cups/d 0.75 (0.64-0.88) ≥5 cups/d 0.77 (0.63-0.92)	Stroke	FFQ (baseline)/ National registries	age; smoking status and pack-years of smoking; education; body mass index; total physical activity; history of diabetes; history of hypertension; aspirin use; family history of myocardial infarction; and intakes of total energy, alcohol, red meat, fish, fruits, and vegetables
Mineharu et	both	13.1	40-79	2012/	men CVD mortality	CVD	FFQ	body mass index (BMI),

al 2011 Japan				76979	<0.14 cup/d 1.00 (1.00-1.00)	mortality	(baseline)/ Mortality certificates at the public health center	history of hypertension, history of diabetes, smoking status, alcohol intake, education, walking hours, hours of sports participation, perceived mental stress, multivitamin use, vitamin E supplement use, consumption of total fruits, total vegetable, total beans, total meat, total fish and seaweeds and total daily energy intake
					0.14-1 cup/d 0.71 (0.53-0.96)			
					1-2 cups/d 0.84 (0.64-0.99)			
					≥3 cups/d 1.17 (0.77-1.76)			
					women CVD mortality			
					<0.14 cup/d 1.00 (1.00-1.00)			
					0.14-1 cup/d 0.87 (0.62-1.23)			
					1-2 cups/d 0.77 (0.55-0.99)			
					≥3 cups/d 2.30 (1.31-4.02)			
Floegel et al 2012 Europe	both	8.9	35-65	704/ 42659	<1 cup/d 1.00 (1.00-1.00)	CVD	FFQ (baseline)/ Confirmed self-reported	age at recruitment, center, sex, smoking, alcohol intake, physical activity, education, employment, vitamin and mineral supplement use during past 4 weeks, total energy intake, tea intake, and decaffeinated coffee intake, BMI, waist-to-hip ratio, and prevalent hypertension
					1-2 cups/d 0.94 (0.64-1.36)			
					2-3 cups/d 1.07 (0.81-1.42)			
					3-4 cups/d 1.02 (0.75-1.38)			
					>4 cups/d 1.10 (0.84-1.44)			
Rautiainen et al 2012 Europe	wo men	9.9	49-83	1114/ 32561	≤2 cups/d 1.00 (1.00-1.00)	CHD	FFQ (baseline)/ National registries	age, education, smoking, body mass index, physical activity, hypertension, hypercholesterolemia,
					3 cups/d 0.87 (0.68-1.12)			
					4 cups/d 0.88 (0.69-1.13)			

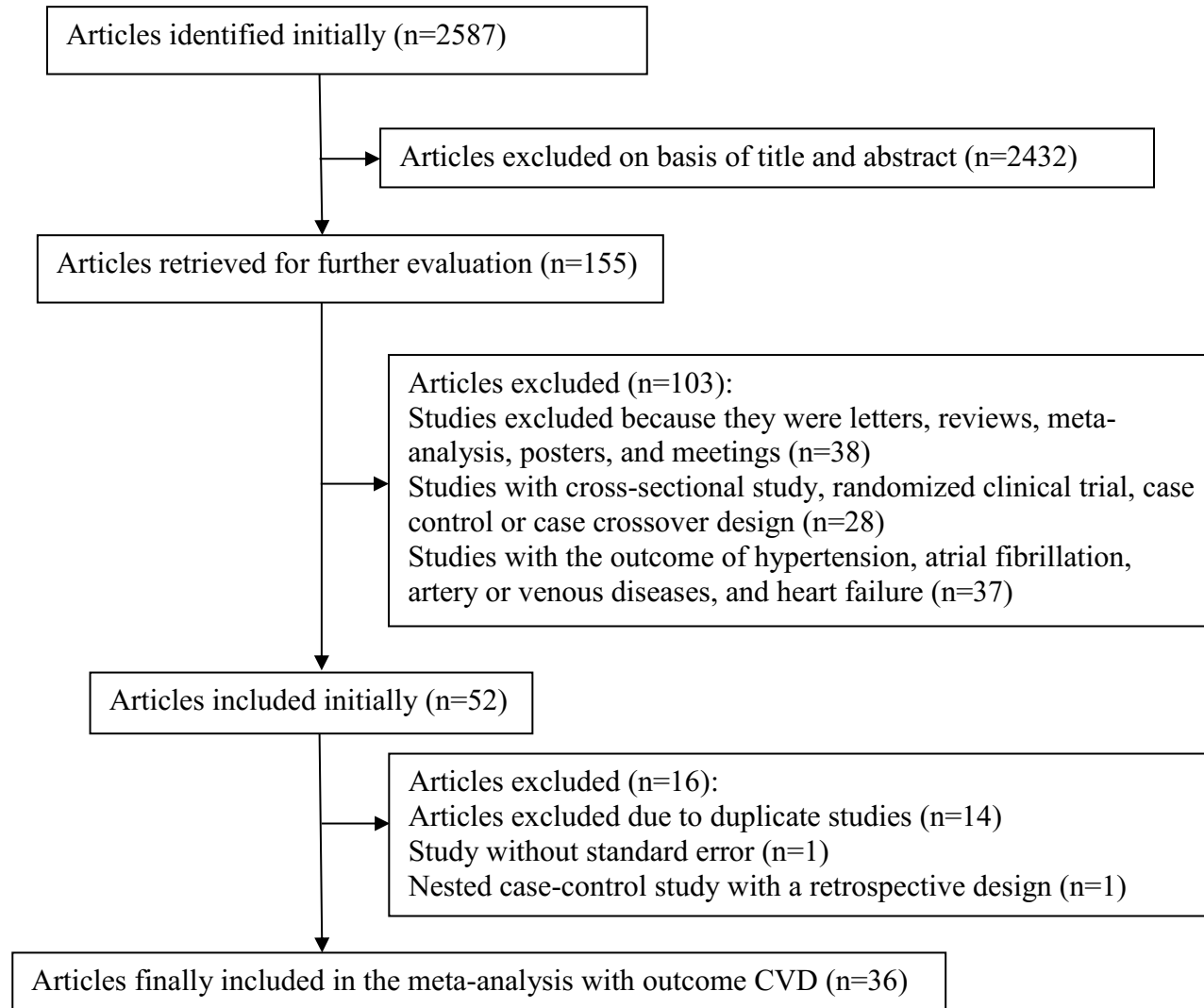
					≥5 cups/d 0.96 (0.72-1.26)			family history of myocardial infarction, aspirin use, hormone replacement therapy use, dietary supplement use, and intakes of total energy and alcohol
Freedman et al 2012 US	both	14	50-71	11828(CHD deaths), 2293(stroke deaths)/ 402260	men CHD mortality	CHD mortality, stroke mortality	FFQ (baseline)/ National registries	age; body-mass index; race or ethnic group; level of education; alcohol consumption; the number of cigarettes smoked per day, use or nonuse of pipes or cigars, and time of smoking cessation; health status; diabetes; marital status; physical activity; total energy intake; consumption of fruits, vegetables, red meat, white meat, and saturated fat; use or nonuse of vitamin supplements; and use or nonuse of postmenopausal hormone therapy
					0 cup/d 1.00 (1.00-1.00)			
					<1 cup/d 0.93 (0.85-1.02)			
					1 cup/d 0.92 (0.84-1.01)			
					2-3 cups/d 0.86 (0.79-0.94)			
					4-5 cups/d 0.87 (0.79-0.96)			
					≥6 cups/d 0.88 (0.78-1.00)			
					women CHD mortality			
					0 cup/d 1.00 (1.00-1.00)			
					<1 cup/d 1.00 (0.89-1.13)			
					1 cup/d 0.91 (0.81-1.03)			
					2-3 cups/d 0.85 (0.76-0.95)			
					4-5 cups/d 0.78 (0.68-0.90)			
					≥6 cups/d 0.72 (0.59-0.88)			
					men stroke mortality			
					0 cup/d 1.00 (1.00-1.00)			
					<1 cup/d 0.99 (0.79-1.24)			

					1 cup/d 0.92 (0.73-1.15)			
					2-3 cups/d 0.84 (0.68-1.02)			
					4-5 cups/d 0.65 (0.51-0.84)			
					≥6 cups/d 0.83 (0.61-1.14)			
					women stroke mortality			
					0 cup/d 1.00 (1.00-1.00)			
					<1 cup/d 1.15 (0.91-1.45)			
					1 cup/d 0.89 (0.70-1.13)			
					2-3 cups/d 0.93 (0.75-1.15)			
					4-5 cups/d 0.82 (0.62-1.09)			
					≥6 cups/d 0.84 (0.56-1.25)			
Kokubo et al 2013 Japan	both	13	45-74	4335/ 82369	Total CVD	CVD, CHD, stroke	FFQ (baseline)/ Confirmed cases	age; sex; smoking; alcohol; body mass index; history of diabetes mellitus; medication of antihypercholesterolemia and antihypertension; sports; dietary intake of fruits, vegetables, fish, and energy; public health centers; and green tea consumption
					0 cup/week 1.00 (1.00-1.00)			
					1-2 cups/week 0.93 (0.86-1.01)			
					3-6 cups/ week 0.89 (0.81- 0.98)			
					1 cups/d 0.84 (0.76-0.92)			
					≥2 cups/d 0.89 (0.80-0.99)			
					Stroke			
					0 cup/week 1.00 (1.00-1.00)			
					1-2 cups/week 0.94 (0.85-1.02)			
					3-6 cups/week 0.89 (0.80-0.99)			
					1 cup/day 0.80 (0.72-0.90)			

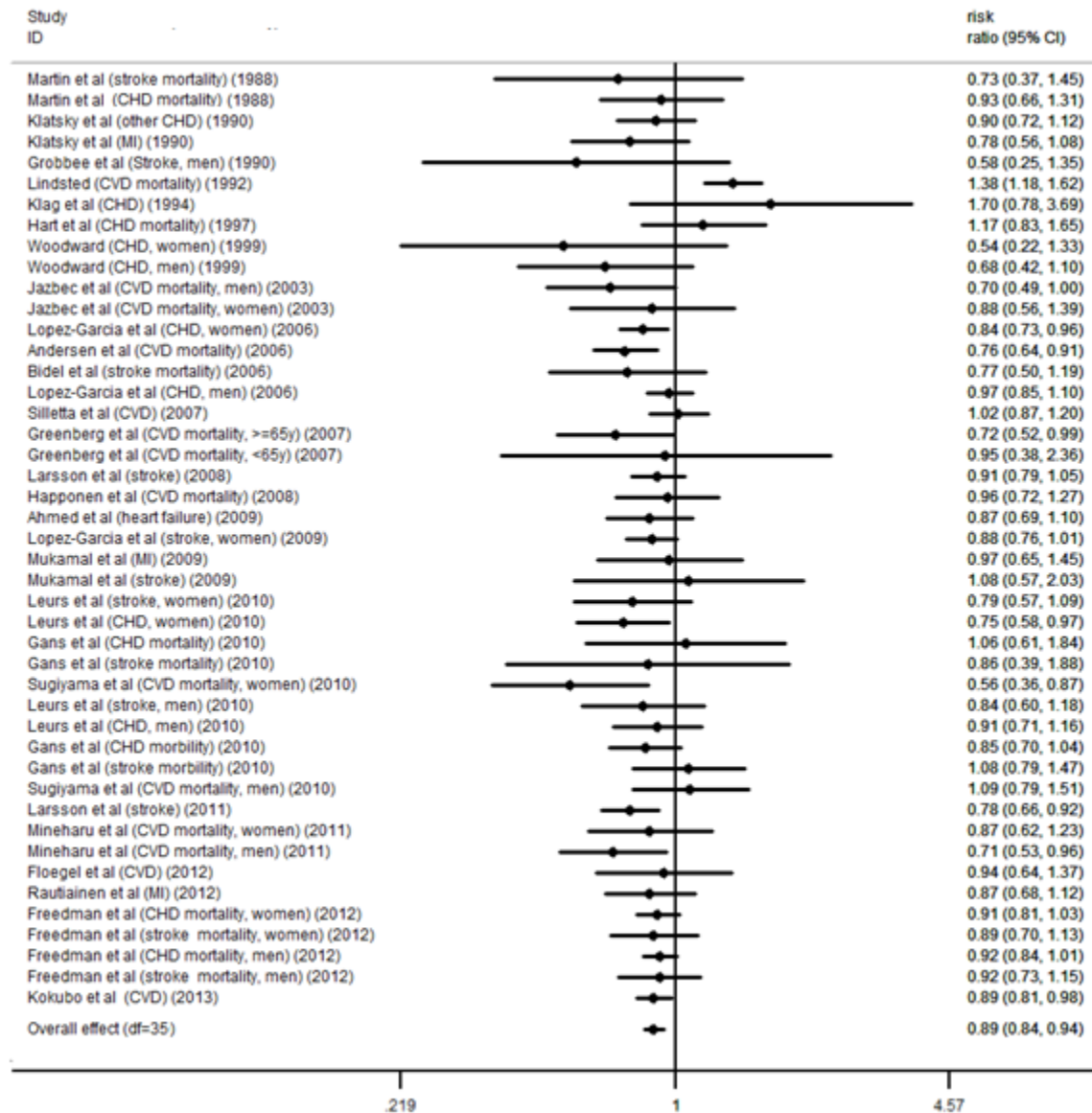


					≥2 cups/day 0.81 (0.72-0.91)			
					CHD			
					0 cup/week 1.00 (1.00-1.00)			
					1-2 cups/week 0.91 (0.76-1.10)			
					3-6 cups/week 0.92 (0.75-1.14)			
					1 cups/d 0.99 (0.81-1.23)			
					≥2 cups/d 1.21 (0.98-1.50)			

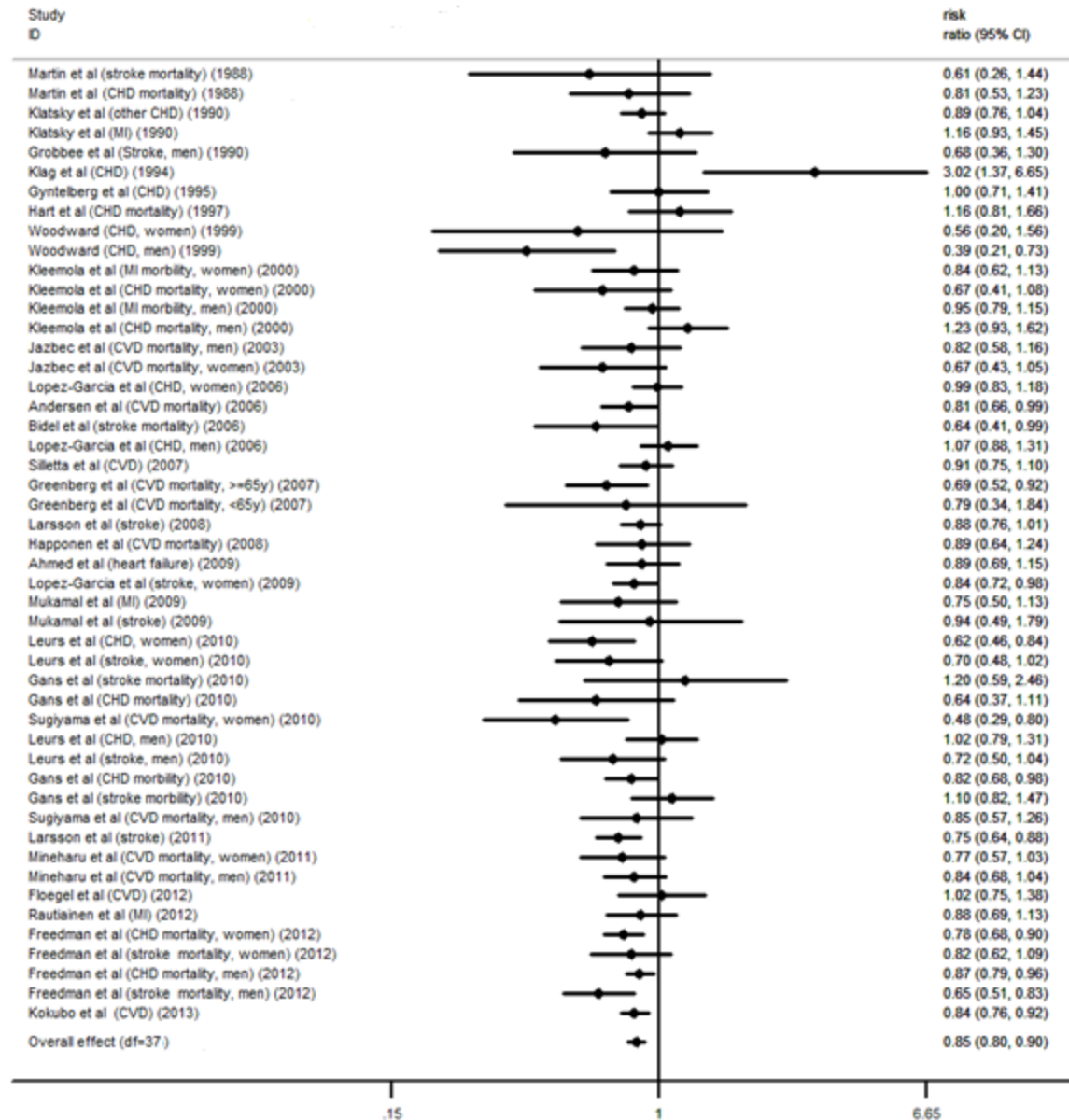
CHD: coronary heart disease; CVD: cardiovascular disease; FFQ: food frequency questionnaire



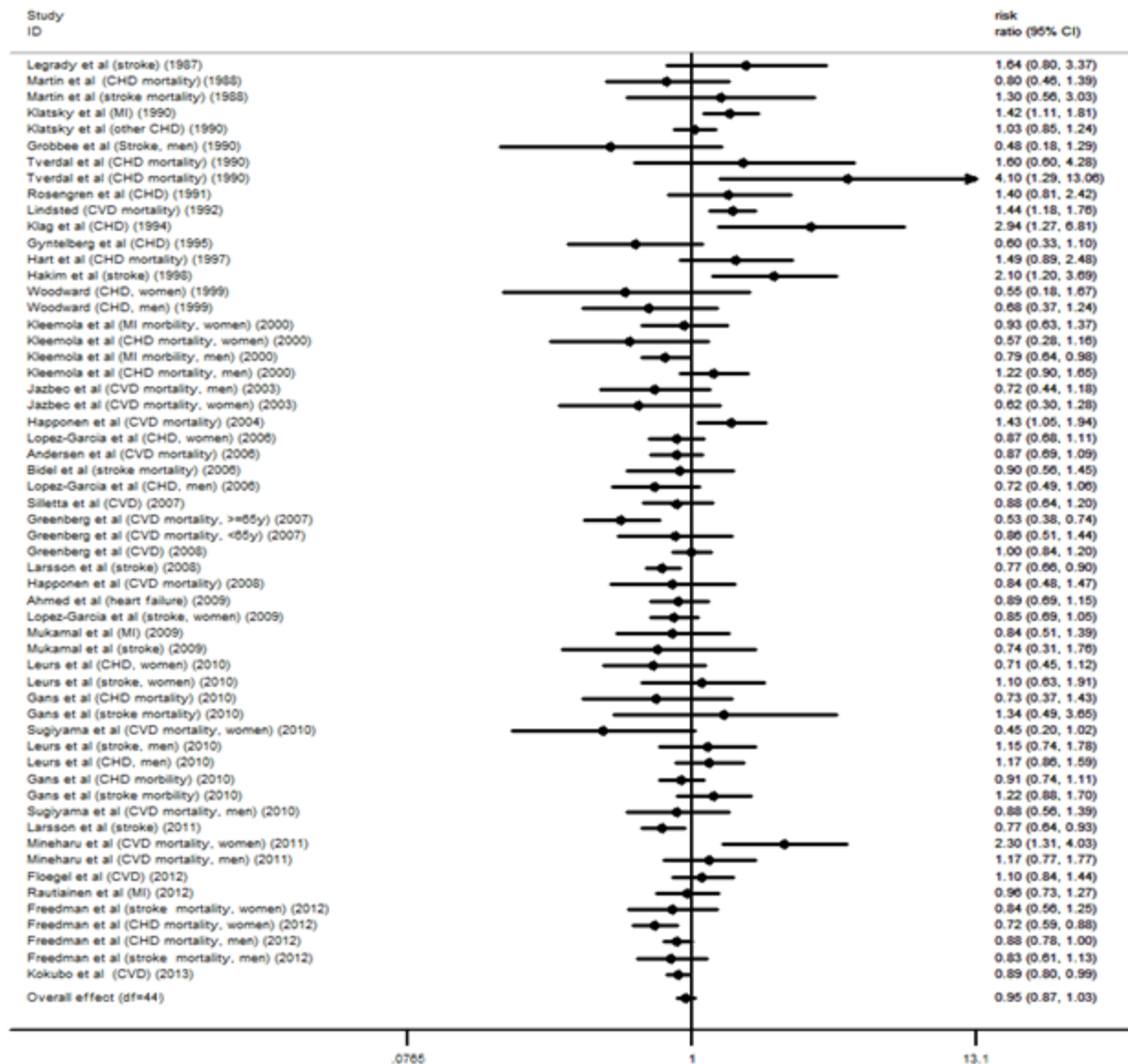
**Figure 1. Study selection process of coffee consumption and risk of CVD**



**Figure 2a. Forest plot of the association between third highest level of coffee consumption (median consumption: 1.5 cups/d; mean: 1.48 cups/d) and risk of CVD compared to the lowest level (median and mean consumption: 0 cup/d). In total 1,279,804 study participants with 36,352 CVD cases were included. The overall effect was obtained from a fixed-effects model that accounted for correlated outcomes. MI means myocardial infarction incidence; CVD means cardiovascular disease incidence; stroke means stroke incidence.**

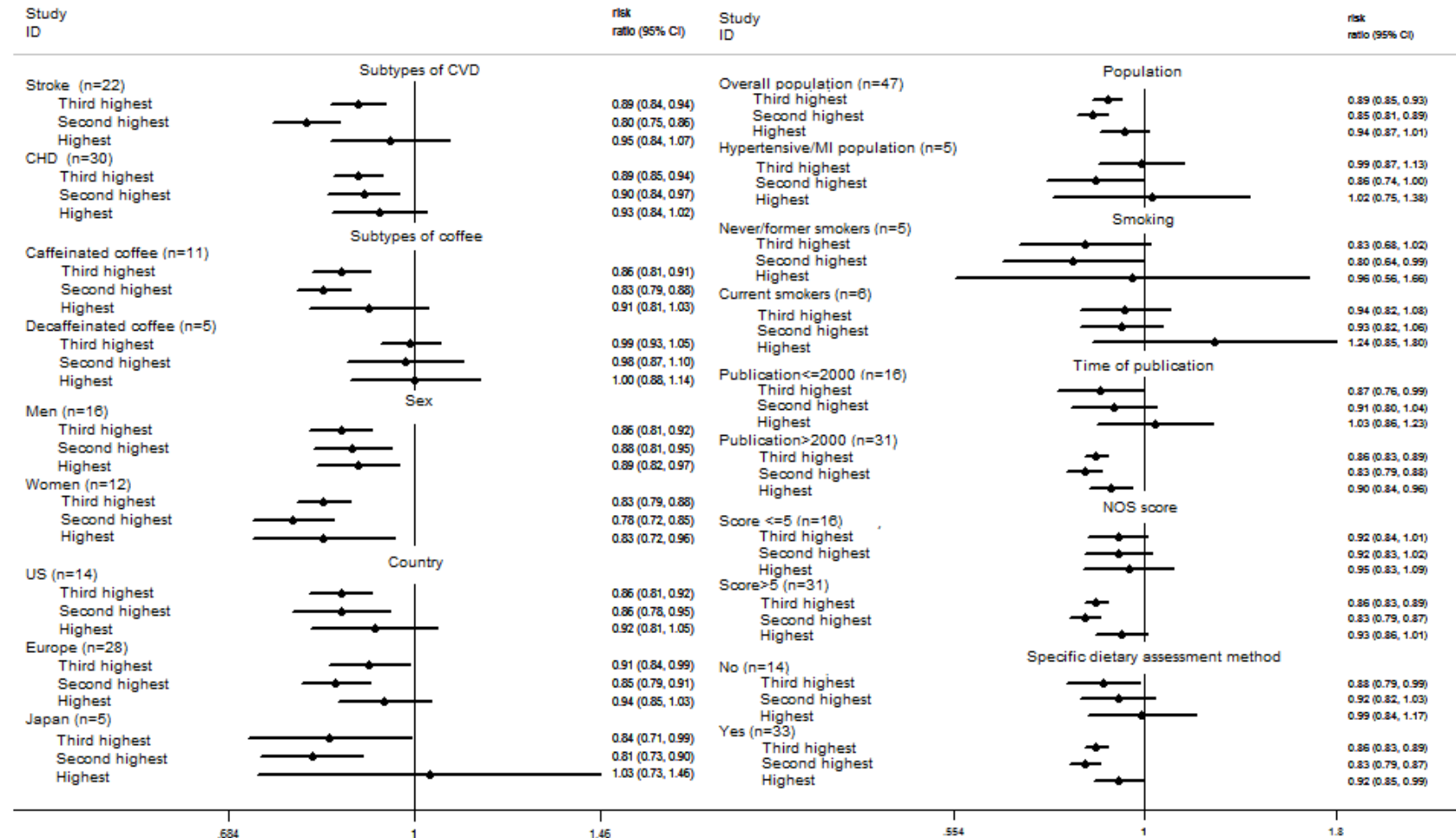


**Figure 2b. Forest plot of the association between second highest level of coffee consumption (median consumption: 3.5 cups/d; mean: 3 cups/d) and risk of CVD compared to the lowest level (median and mean consumption: 0 cups/d). In total 1,279,804 study participants with 36,352 CVD cases were included. The overall effect was obtained from a fixed-effects model that accounted for correlated outcomes. MI means myocardial infarction incidence; CVD means cardiovascular disease incidence; stroke means stroke incidence.**

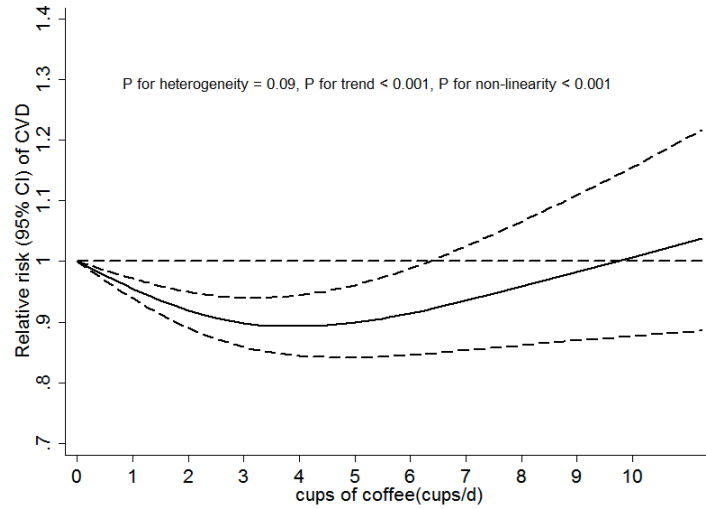


**Figure 2c. Forest plot of the association between highest level of coffee consumption (median consumption: 5 cups/d; mean: 5.5 cups/d) and risk of CVD compared to the lowest level (median and mean consumption: 0 cups/d). In total 1,279,804 study participants with 36,352 CVD cases were included. The overall effect was obtained from a fixed-effects model that accounted for correlated outcomes. MI means myocardial infarction incidence; CVD means cardiovascular disease incidence; stroke means stroke incidence.**

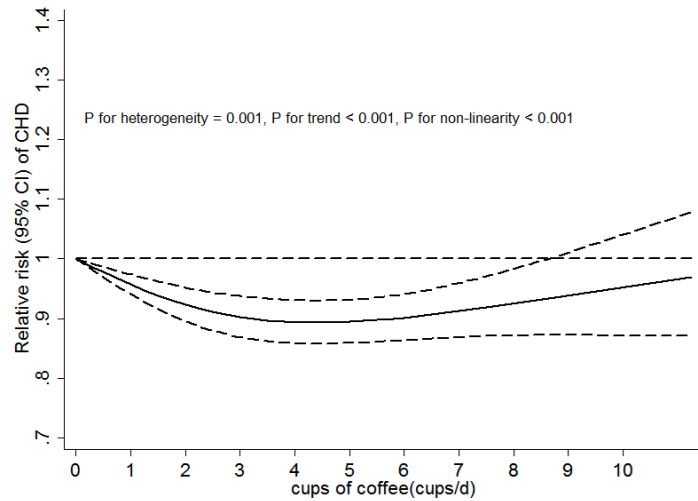




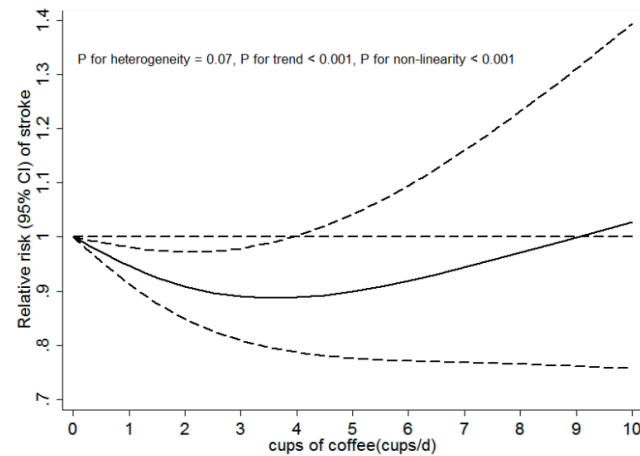
**Figure 3. Stratified analysis of the association between coffee consumption and risk of CVD.** The included studies for the stratified analysis were the same as that for the dose response analysis. n was the number of comparisons for the highest level of coffee consumption. NOS score: the score using the Newcastle–Ottawa scale; specific dietary assessment method: diet that was assessed by 24h diet recall, diet record or food frequency questionnaire.



a. Coffee consumption and risk of CVD (n = 47)



b. Coffee consumption and risk of CHD (n = 31)



c. Coffee consumption and risk of stroke (n = 22)

**Figure 4. Dose response relationships of coffee consumption with risk of CVD. n was the number of comparisons.**

**Supplemental table 1: the quality assessment of included studies using the Newcastle–Ottawa scale <sup>1</sup>**

[illegible]

al											
Bidel et al	2006	1	1	0	1	1	1	1	1	0	7
Silletta et al	2007	0	1	0	0	0	1	1	0	0	3
Greenberg et al	2007	1	1	0	1	1	0	1	1	0	6
Happonen et al	2008	1	1	0	0	0	0	1	1	1	5
Greenberg et al	2008	1	1	1	0	0	1	1	1	0	6
Larsson et al	2008	1	1	1	0	1	1	1	1	0	7
Mukamal et al	2009	1	1	0	0	0	0	1	1	0	4
Ahmed et al	2009	1	1	1	1	0	0	1	1	0	6
Lopez-Garcia et al	2009	0	1	1	1	1	0	1	1	0	6
Gans et al	2010	1	1	1	1	0	0	1	1	0	6
Leurs et al	2010	1	1	1	1	1	0	1	1	0	7
Sugiyama et al	2010	1	1	1	1	1	1	1	1	0	8
Mineharu et al	2011	1	1	1	1	1	1	1	1	0	8
Larsson et al	2011	1	1	1	1	1	1	1	1	1	9
Freedman et al	2012	0	1	1	1	1	0	1	1	0	6
Rautiainen et al	2012	1	1	1	1	0	1	1	1	1	8
Floegel et al	2012	1	1	0	1	1	1	1	1	0	7
Kokubo et al	2013	1	1	1	1	1	1	1	1	1	9

The quality of included studies was assessed by the Newcastle Ottawa scale. A study can be awarded a maximum of one star for each numbered item within the Selection and Outcome categories and a maximum of two stars for Comparability.

**Selection:** 1) Representativeness of exposed cohort: 1, study population truly or somewhat representative of a community/ population based study; 0, study population was sampled from a special population, ie. population from a company, hospital patients, data from the health insurance company or health examination organization, nurses, Adventist group.

2) Selection of non-exposed cohort: 1, drawn from the same community as the exposed cohort.

3) Ascertainment of exposure: 1, specific dietary assessment method of coffee consumption (FFQ/diet record/24h diet recall) with validation; 0, no specific dietary assessment method or specific dietary assessment method without validation

4) Demonstration that outcome was not present at start of study: 1, exclusion of participants with a history of CVD at the beginning of the study.

**Comparability:** 1) 1, whether a study adjusted for smoking deliberately (not only adjust for the smoking status, but also the number of cigarettes or duration of smoking); 1, whether a study adjusted for baseline hypertension.

**Outcome:** 1) Assessment of outcome: 1, CVD cases were confirmed by medical records or record linkage; 0, self-reported.

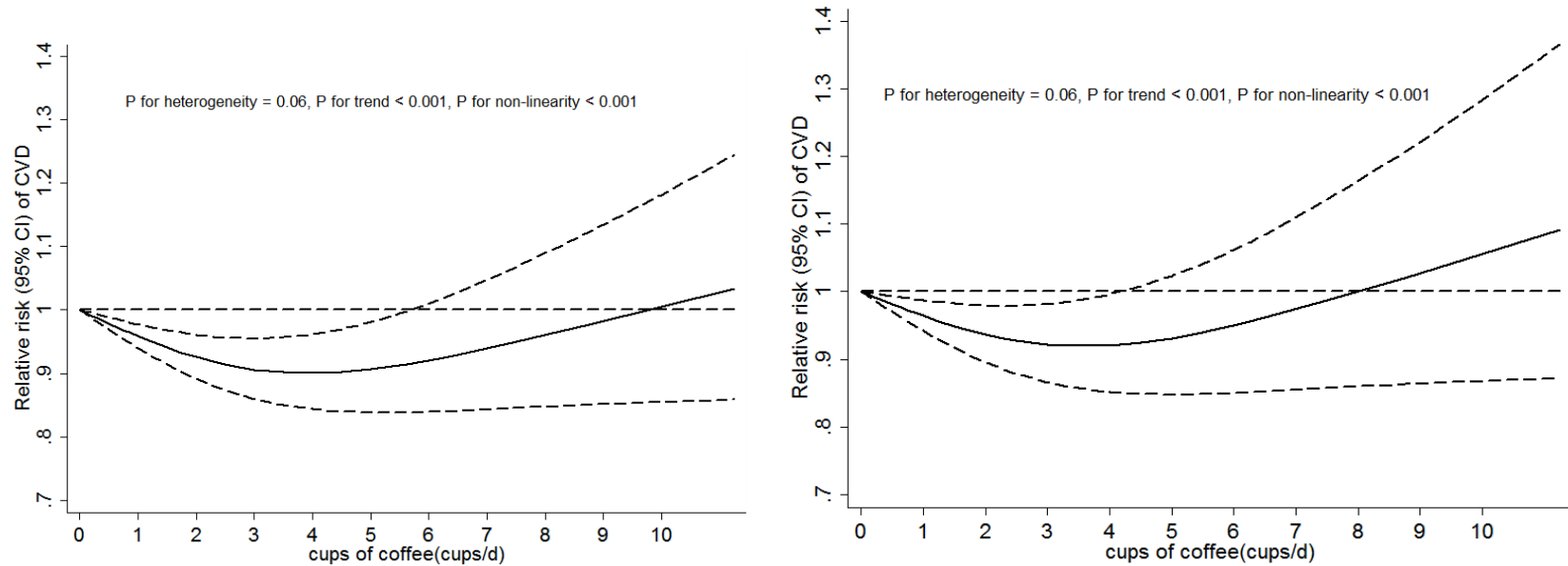
2) Was follow-up long enough for outcomes to occur: 1, duration of follow-up  $\geq 5$  year; 0, if duration of follow-up  $< 5$  year.

3) Loss to follow-up rate: 1, complete follow-up or loss to follow up rate  $\leq 20\%$ ; 0, follow-up rate  $< 80\%$  or no description of those lost.

**Supplemental table 2: Egger's test for the publication bias on coffee consumption and risk of type 2 diabetes**

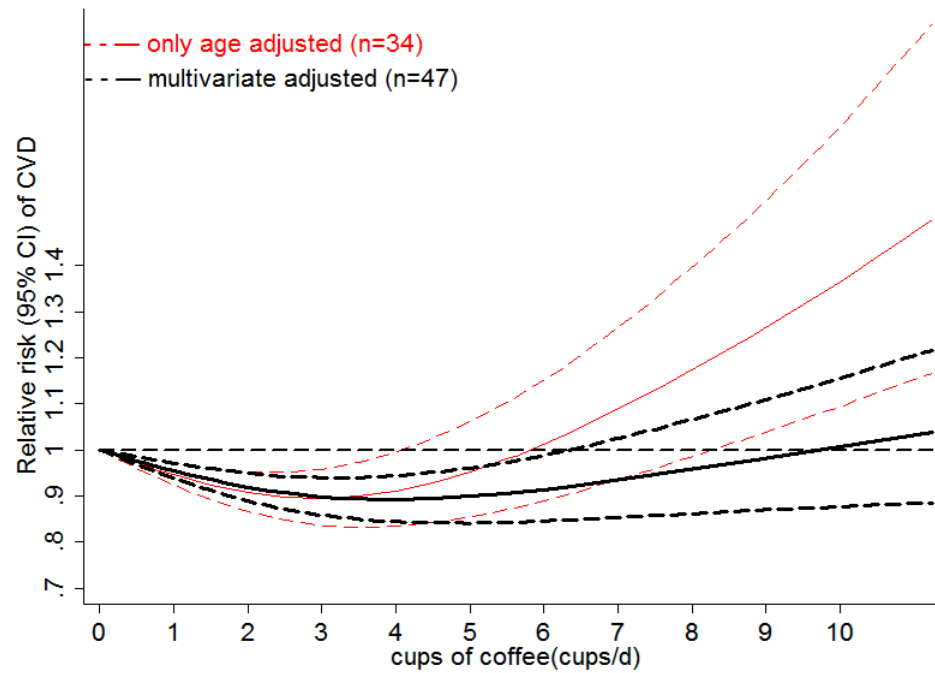
P value of Egger's test	Total publications
Highest	0.28
Second highest	0.42
Third highest	0.32

## Supplemental figures



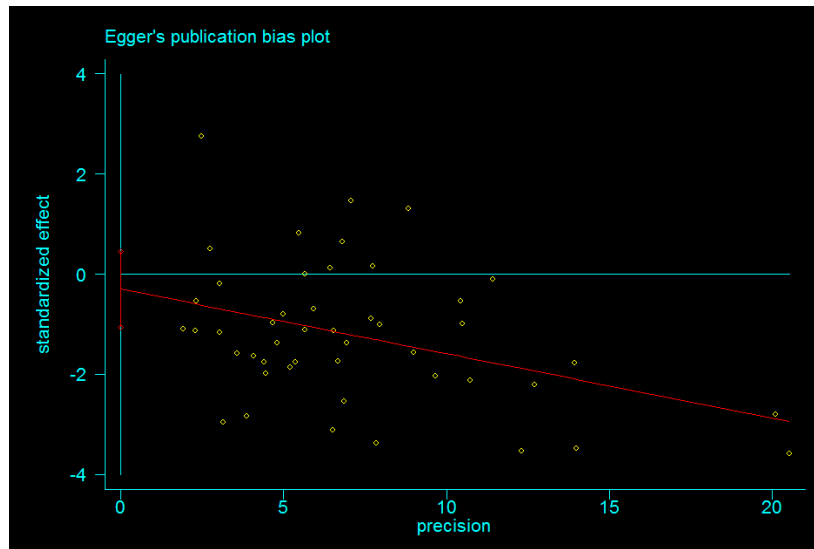
a. Coffee consumption and risk of CVD choosing CHD for correlated outcomes (n=36) b. Coffee consumption and risk of CVD choosing stroke for correlated outcomes (n=37)

**Supplemental figure 1: Dose response relationship of coffee consumption with cardiovascular disease risk choosing only one outcome for correlated outcomes within the same study. n was the number of comparisons.**

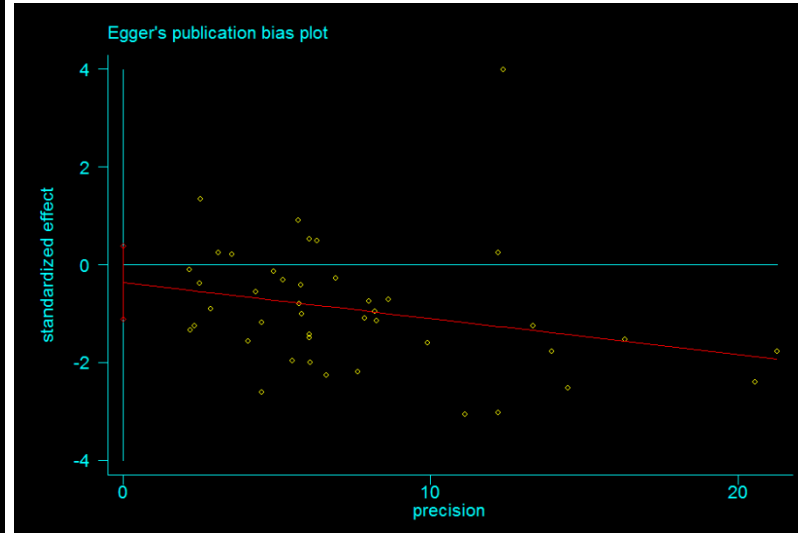


**Supplemental figure 2. Dose response relationship of coffee consumption with cardiovascular disease risk from models adjusted for different confounders. Red curve: included studies only adjusted for age; Black curve: included studies with multivariate adjusted models**

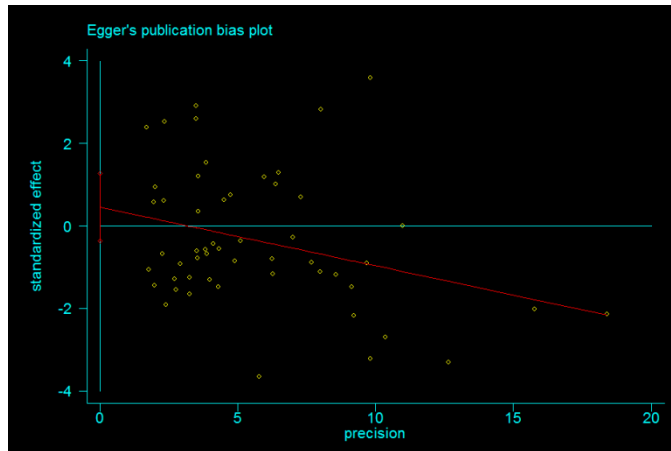




a. The third highest category of coffee consumption



b. The second highest category of coffee consumption



c. The highest category of coffee consumption

**Supplemental figure 3: Egger's test for publication bias for the association between coffee consumption and risk of CVD**

The study selection process:

Of the 53 initially included studies, we excluded 14 studies due to duplicate publication<sup>2-15</sup>, one study with point estimate without standard error<sup>16</sup>, and one nested case control study with a retrospective design<sup>17</sup>. Thirty-six studies were remained in the meta-analysis<sup>18-53</sup>.

### Reference

1. Milionis HJ, Kostapanos MS, Liberopoulos EN, Goudevenos J, Athyros VG, Mikhailidis DP, Elisaf MS. Different definitions of the metabolic syndrome and risk of first-ever acute ischaemic non-embolic stroke in elderly subjects. *International journal of clinical practice*. 2007;61:545-551
2. Dawber TR, Kannel WB, Gordon T. Coffee and cardiovascular disease. Observations from the framingham study. *The New England journal of medicine*. 1974;291:871-874
3. LaCroix AZ, Mead LA, Liang KY, Thomas CB, Pearson TA. Coffee consumption and the incidence of coronary heart disease. *The New England journal of medicine*. 1986;315:977-982
4. Willett WC, Stampfer MJ, Manson JE, Colditz GA, Rosner BA, Speizer FE, Hennekens CH. Coffee consumption and coronary heart disease in women. A ten-year follow-up. *JAMA : the journal of the American Medical Association*. 1996;275:458-462
5. Rosner SA, Akesson A, Stampfer MJ, Wolk A. Coffee consumption and risk of myocardial infarction among older swedish women. *American journal of epidemiology*. 2007;165:288-293
6. Zhang W, Lopez-Garcia E, Li TY, Hu FB, van Dam RM. Coffee consumption and risk of cardiovascular diseases and all-cause mortality among men with type 2 diabetes. *Diabetes care*. 2009;32:1043-1045
7. Zhang WL, Lopez-Garcia E, Li TY, Hu FB, van Dam RM. Coffee consumption and risk of cardiovascular events and all-cause mortality among women with type 2 diabetes. *Diabetologia*. 2009;52:810-817
8. Levitan EB, Ahmed HN, Mittleman MA, Wolk A. Coffee consumption and incidence of heart failure in women. *Circulation. Heart failure*. 2011;4:414-418
9. Wang Y, Tuomilehto J, Jousilahti P, Antikainen R, Mahonen M, Mannisto S, Katzmarzyk PT, Hu G. Coffee consumption and the risk of heart failure in finnish men and women. *Heart*. 2011;97:44-48
10. Mukamal KJ, Maclure M, Muller JE, Sherwood JB, Mittleman MA. Caffeinated coffee consumption and mortality after acute myocardial infarction. *American heart journal*. 2004;147:999-1004
11. Lopez-Garcia E, van Dam RM, Li TY, Rodriguez-Artalejo F, Hu FB. The relationship of coffee consumption with mortality. *Annals of internal medicine*. 2008;148:904-914
12. Lopez-Garcia E, Rodriguez-Artalejo F, Li TY, Mukamal KJ, Hu FB, van Dam RM. Coffee consumption and mortality in women with cardiovascular disease. *The American journal of clinical nutrition*. 2011;94:218-224

13. Stensvold I, Tverdal A. The relationship of coffee consumption to various self-reported cardiovascular events in middle-aged norwegian men and women. *Scandinavian journal of social medicine*. 1995;23:103-109
14. Klatsky AL, Armstrong MA, Friedman GD. Coffee, tea, and mortality. *Annals of epidemiology*. 1993;3:375-381
15. Wilhelmsen L, Rosengren A, Eriksson H, Lappas G. Heart failure in the general population of men--morbidity, risk factors and prognosis. *Journal of internal medicine*. 2001;249:253-261
16. Murray SS, Bjelke E, Gibson RW, Schuman LM. Coffee consumption and mortality from ischemic heart disease and other causes: Results from the lutheran brotherhood study, 1966-1978. *American journal of epidemiology*. 1981;113:661-667
17. Nilsson LM, Wennberg M, Lindahl B, Eliasson M, Jansson JH, Van Guelpen B. Consumption of filtered and boiled coffee and the risk of first acute myocardial infarction; a nested case/referent study. *Nutrition, metabolism, and cardiovascular diseases : NMCD*. 2010;20:527-535
18. LeGrady D, Dyer AR, Shekelle RB, Stamler J, Liu K, Paul O, Lepper M, Shryock AM. Coffee consumption and mortality in the chicago western electric company study. *American journal of epidemiology*. 1987;126:803-812
19. Martin JB, Annegers JF, Curb JD, Heyden S, Howson C, Lee ES, Lee M. Mortality patterns among hypertensives by reported level of caffeine consumption. *Preventive medicine*. 1988;17:310-320
20. Grobbee DE, Rimm EB, Giovannucci E, Colditz G, Stampfer M, Willett W. Coffee, caffeine, and cardiovascular disease in men. *The New England journal of medicine*. 1990;323:1026-1032
21. Tverdal A, Stensvold I, Solvoll K, Foss OP, Lund-Larsen P, Bjartveit K. Coffee consumption and death from coronary heart disease in middle aged norwegian men and women. *Bmj*. 1990;300:566-569
22. Rosengren A, Wilhelmsen L. Coffee, coronary heart disease and mortality in middle-aged swedish men: Findings from the primary prevention study. *Journal of internal medicine*. 1991;230:67-71
23. Klag MJ, Mead LA, LaCroix AZ, Wang NY, Coresh J, Liang KY, Pearson TA, Levine DM. Coffee intake and coronary heart disease. *Annals of epidemiology*. 1994;4:425-433
24. Gyntelberg F, Hein HO, Suadicani P, Sorensen H. Coffee consumption and risk of ischaemic heart disease--a settled issue? *Journal of internal medicine*. 1995;237:55-61
25. Hart C, Smith GD. Coffee consumption and coronary heart disease mortality in scottish men: A 21 year follow up study. *Journal of epidemiology and community health*. 1997;51:461-462
26. Hakim AA, Ross GW, Curb JD, Rodriguez BL, Burchfiel CM, Sharp DS, Yano K, Abbott RD. Coffee consumption in hypertensive men in older middle-age and the risk of stroke: The honolulu heart program. *Journal of clinical epidemiology*. 1998;51:487-494
27. Woodward M, Tunstall-Pedoe H. Coffee and tea consumption in the scottish heart health study follow up: Conflicting relations with coronary risk factors, coronary disease, and all cause mortality. *Journal of epidemiology and community health*. 1999;53:481-487
28. Kleemola P, Jousilahti P, Pietinen P, Vartiainen E, Tuomilehto J. Coffee consumption and the risk of coronary heart disease and death. *Archives of internal medicine*. 2000;160:3393-3400
29. Jazbec A, Simic D, Corovic N, Durakovic Z, Pavlovic M. Impact of coffee and other selected factors on general mortality and mortality due to cardiovascular disease in croatia. *Journal of health, population, and nutrition*. 2003;21:332-340

30. Happonen P, Voutilainen S, Salonen JT. Coffee drinking is dose-dependently related to the risk of acute coronary events in middle-aged men. *The Journal of nutrition*. 2004;134:2381-2386
31. Lopez-Garcia E, van Dam RM, Willett WC, Rimm EB, Manson JE, Stampfer MJ, Rexrode KM, Hu FB. Coffee consumption and coronary heart disease in men and women: A prospective cohort study. *Circulation*. 2006;113:2045-2053
32. Andersen LF, Jacobs DR, Jr., Carlsen MH, Blomhoff R. Consumption of coffee is associated with reduced risk of death attributed to inflammatory and cardiovascular diseases in the iowa women's health study. *The American journal of clinical nutrition*. 2006;83:1039-1046
33. Silletta MG, Marfisi R, Levantesi G, Boccanelli A, Chieffo C, Franzosi M, Geraci E, Maggioni AP, Nicolosi G, Schweiger C, Tavazzi L, Tognoni G, Marchioli R, Investigators GI-P. Coffee consumption and risk of cardiovascular events after acute myocardial infarction: Results from the gissi (gruppo italiano per lo studio della sopravvivenza nell'infarto miocardico)-prevenzione trial. *Circulation*. 2007;116:2944-2951
34. Greenberg JA, Chow G, Ziegelstein RC. Caffeinated coffee consumption, cardiovascular disease, and heart valve disease in the elderly (from the framingham study). *The American journal of cardiology*. 2008;102:1502-1508
35. Happonen P, Laara E, Hiltunen L, Luukinen H. Coffee consumption and mortality in a 14-year follow-up of an elderly northern finnish population. *The British journal of nutrition*. 2008;99:1354-1361
36. Larsson SC, Mannisto S, Virtanen MJ, Kontto J, Albanes D, Virtamo J. Coffee and tea consumption and risk of stroke subtypes in male smokers. *Stroke; a journal of cerebral circulation*. 2008;39:1681-1687
37. Mukamal KJ, Hallqvist J, Hammar N, Ljung R, Gemes K, Ahlbom A, Ahnve S, Janszky I. Coffee consumption and mortality after acute myocardial infarction: The stockholm heart epidemiology program. *American heart journal*. 2009;157:495-501
38. Ahmed HN, Levitan EB, Wolk A, Mittleman MA. Coffee consumption and risk of heart failure in men: An analysis from the cohort of swedish men. *American heart journal*. 2009;158:667-672
39. Lopez-Garcia E, Rodriguez-Artalejo F, Rexrode KM, Logroscino G, Hu FB, van Dam RM. Coffee consumption and risk of stroke in women. *Circulation*. 2009;119:1116-1123
40. Leurs LJ, Schouten LJ, Goldbohm RA, van den Brandt PA. Total fluid and specific beverage intake and mortality due to ihd and stroke in the netherlands cohort study. *The British journal of nutrition*. 2010;104:1212-1221
41. Mineharu Y, Koizumi A, Wada Y, Iso H, Watanabe Y, Date C, Yamamoto A, Kikuchi S, Inaba Y, Toyoshima H, Kondo T, Tamakoshi A, Group Js. Coffee, green tea, black tea and oolong tea consumption and risk of mortality from cardiovascular disease in japanese men and women. *Journal of epidemiology and community health*. 2011;65:230-240
42. Rautiainen S, Levitan EB, Orsini N, Akesson A, Morgenstern R, Mittleman MA, Wolk A. Total antioxidant capacity from diet and risk of myocardial infarction: A prospective cohort of women. *The American journal of medicine*. 2012;125:974-980
43. Wilhelmsen L, Tibblin G, Elmfeldt D, Wedel H, Werko L. Coffee consumption and coronary heart disease in middle-aged swedish men. *Acta medica Scandinavica*. 1977;201:547-552
44. Klatsky AL, Friedman GD, Armstrong MA. Coffee use prior to myocardial infarction restudied: Heavier intake may increase the risk. *American journal of epidemiology*. 1990;132:479-488

45. de Koning Gans JM, Uiterwaal CS, van der Schouw YT, Boer JM, Grobbee DE, Verschuren WM, Beulens JW. Tea and coffee consumption and cardiovascular morbidity and mortality. *Arteriosclerosis, thrombosis, and vascular biology*. 2010;30:1665-1671
46. Floegel A, Pischon T, Bergmann MM, Teucher B, Kaaks R, Boeing H. Coffee consumption and risk of chronic disease in the european prospective investigation into cancer and nutrition (epic)-germany study. *The American journal of clinical nutrition*. 2012;95:901-908
47. Lindsted KD, Kuzma JW, Anderson JL. Coffee consumption and cause-specific mortality. Association with age at death and compression of mortality. *Journal of clinical epidemiology*. 1992;45:733-742
48. Sugiyama K, Kuriyama S, Akhter M, Kakizaki M, Nakaya N, Ohmori-Matsuda K, Shimazu T, Nagai M, Sugawara Y, Hozawa A, Fukao A, Tsuji I. Coffee consumption and mortality due to all causes, cardiovascular disease, and cancer in japanese women. *The Journal of nutrition*. 2010;140:1007-1013
49. Freedman ND, Park Y, Abnet CC, Hollenbeck AR, Sinha R. Association of coffee drinking with total and cause-specific mortality. *The New England journal of medicine*. 2012;366:1891-1904
50. Kokubo Y, Iso H, Saito I, Yamagishi K, Yatsuya H, Ishihara J, Inoue M, Tsugane S. The impact of green tea and coffee consumption on the reduced risk of stroke incidence in japanese population: The japan public health center-based study cohort. *Stroke; a journal of cerebral circulation*. 2013;44:1369-1374
51. Bidel S, Hu G, Qiao Q, Jousilahti P, Antikainen R, Tuomilehto J. Coffee consumption and risk of total and cardiovascular mortality among patients with type 2 diabetes. *Diabetologia*. 2006;49:2618-2626
52. Greenberg JA, Dunbar CC, Schnoll R, Kokolis R, Kokolis S, Kassotis J. Caffeinated beverage intake and the risk of heart disease mortality in the elderly: A prospective analysis. *The American journal of clinical nutrition*. 2007;85:392-398
53. Larsson SC, Virtamo J, Wolk A. Coffee consumption and risk of stroke in women. *Stroke; a journal of cerebral circulation*. 2011;42:908-912

## CHAPTER 2

---

# ASSOCIATION OF COFFEE CONSUMPTION WITH TOTAL AND CAUSE-SPECIFIC MORTALITY IN THREE LARGE PROSPECTIVE COHORTS

---

Ming Ding, MD <sup>1</sup>; Ambika Satija, BA <sup>1</sup>; Shilpa N Bhupathiraju, PhD <sup>1</sup>; Yang Hu, MS <sup>1</sup>; Qi Sun, MD, DSc <sup>1,3</sup>; Jiali Han, DSc <sup>5,6</sup>; Esther Lopez-Garcia, PhD<sup>7</sup>; Walter Willett, MD, DrPH <sup>1,2,3</sup>; Rob M. van Dam, PhD <sup>1,8</sup>; Frank B. Hu, MD, PhD <sup>1,2,3</sup>

<sup>1</sup>Department of Nutrition, Harvard School of Public Health, Boston, MA

<sup>2</sup>Department of Epidemiology, Harvard School of Public Health, Boston, MA

<sup>3</sup>Channing Division of Network Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA

<sup>5</sup>Department of Epidemiology, Fairbanks School of Public Health, Indiana University, Indianapolis, Indiana

<sup>6</sup>Melvin and Bren Simon Cancer Center, Indiana University, Indianapolis, Indiana

<sup>7</sup>Department of Preventive Medicine and Public Health, School of Medicine, Universidad Autónoma de Madrid/IdiPAZ, CIBER of Epidemiology and Public Health (CIBERESP), Spain

<sup>8</sup>Saw Swee Hock School of Public Health and Yong Loo Lin School of Medicine, National University of Singapore and National University Health System, Singapore

## ABSTRACT

**Background-**The association between consumption of caffeinated and decaffeinated coffee and risk of mortality remains inconclusive.

**Methods and Results-**We examined the associations of consumption of total, caffeinated, and decaffeinated coffee with risk of subsequent total and cause-specific mortality among 74,890 women in the Nurses' Health Study (NHS), 93,054 women in the NHS 2, and 40,557 men in the Health Professionals Follow-up Study. Coffee consumption was assessed at baseline using a semi-quantitative food frequency questionnaire. During 4,690,072 person-years of follow-up, 19,524 women and 12,432 men died. Consumption of total, caffeinated, and decaffeinated coffee were non-linearly associated with mortality. Compared to non-drinkers, coffee consumption one to five cups/d was associated with lower risk of mortality, while coffee consumption more than five cups/d was not associated with risk of mortality. However, when restricting to never smokers, compared to non-drinkers, the HRs of mortality were 0.94 (0.89 to 0.99) for  $\leq 1$  cup/d, 0.92 (0.87 to 0.97) for 1.1-3 cups/d, 0.85 (0.79 to 0.92) for 3.1-5 cups/d, and 0.88 (0.78 to 0.99) for  $> 5$  cups/d ( $p$  for non-linearity = 0.32;  $p$  for trend  $< 0.001$ ). Significant inverse associations were observed for caffeinated ( $p$  for trend  $< 0.001$ ) and decaffeinated coffee ( $p$  for trend = 0.022). Significant inverse associations were observed between coffee consumption and deaths due to cardiovascular disease, neurological diseases, and suicide. No significant association between coffee consumption and total cancer mortality was found.

**Conclusions-**Higher consumption of total coffee, caffeinated coffee, and decaffeinated coffee was associated with lower risk of total mortality.

**Key Words:** caffeinated coffee, decaffeinated coffee, smoking, mortality

## INTRODUCTION

Coffee is one of the most commonly consumed beverages worldwide. The associations between coffee consumption and risks of several disease outcomes have been investigated. Coffee consumption has been inversely associated with risks of type 2 diabetes <sup>1</sup>, liver cancer <sup>2</sup>, endometrial cancer <sup>3</sup>, lethal prostate cancer <sup>4</sup>, basal cell carcinoma of the skin <sup>5</sup>, and neurological diseases <sup>6</sup>, as well as with risk of cardiovascular disease (CVD) when consumed in moderation <sup>7</sup>.

The association between coffee consumption and risk of total mortality has also been investigated. Recent studies showed an inverse association between moderate coffee consumption and risk of mortality, and an inverse or null association between heavy coffee consumption and risk of mortality <sup>8-16</sup>. However, some earlier studies <sup>17-19</sup> and a recent study <sup>20</sup> found heavy coffee consumption to be associated with higher risk of mortality. Summarizing individual studies, meta-analyses have concluded that coffee consumption is not associated with higher risk of mortality; however, there was significant between-study heterogeneity in the effect estimates <sup>21-23</sup>. The associations between coffee consumption and cause-specific mortality, especially CVD and cancer mortality, have been sporadically investigated together with total mortality, <sup>8, 10, 11, 23</sup> with most studies finding an inverse association with CVD mortality, and no association with cancer mortality.

Based on the results of previous studies, four questions remain unanswered: first, does a non-linear relationship exist between coffee consumption and risk of mortality, i.e., is moderate coffee consumption associated with lower risk of mortality and heavy coffee drinking not associated with risk of mortality or even with an increased risk? Second, if a non-linear association exists, is it truly a biological effect of coffee or is it an artifact due to the confounding of smoking? Third, what are the associations of coffee consumption with risks of cause-specific



mortality? Fourth, do caffeinated and decaffeinated coffee have similar associations with risk of mortality?

We therefore examined the association of coffee consumption with total and cause-specific mortality in three large, ongoing, independent cohort studies of men and women. This analysis updated our earlier publication on coffee consumption and total mortality in the Nurses' Health Study (NHS) and Health Professionals Follow-up Study (HPFS) with 6,888 total deaths and extended to a younger cohort of nurses (Nurses' Health Study II). These cohorts provide measures of caffeinated and decaffeinated coffee consumption, extensive data on known or suspected confounders, and up to 30 years of follow-up, during which more than 30,000 deaths have been recorded.

## **METHODS**

### **Study Population**

The NHS began in 1976, when 121,700 female registered nurses aged 30-55 y residing in 11 states were recruited to complete a baseline questionnaire about their lifestyle and medical history. The NHS II was established in 1989 and consisted of 116,671 younger female registered nurses, aged 25-42 y at baseline. These nurses responded to a baseline questionnaire similar to the NHS. The HPFS was initiated in 1986, and was composed of 51,529 male dentists, pharmacists, veterinarians, optometrists, osteopathic physicians, and podiatrists, aged 40-75 y at baseline. The male participants returned a baseline questionnaire about detailed medical history, lifestyle, and usual diet. In all three cohorts, questionnaires were collected at baseline and biennially thereafter, to update information on lifestyle factors and the occurrence of chronic diseases. All of the three cohorts consist of approximate 95% Caucasians.

For the current analysis, we excluded participants who reported CVD, or cancer at baseline (1984 for the NHS, 1991 for the NHS II, and 1986 for the HPFS). We further excluded participants with missing caffeinated or decaffeinated coffee consumption at baseline, those who left more than 70 food items blank or had daily energy intakes  $< 600$  or  $> 3500$  kcal for women and  $< 800$  or  $> 4200$  kcal for men. The study protocol was approved by the institutional review boards of Brigham and Women's Hospital and Harvard School of Public Health.

### **Assessment of Coffee Consumption**

In 1984, a 116-item food frequency questionnaire (FFQ) was administered to the NHS participants to obtain information on usual intake of food and beverages. Starting in 1986, an expanded 131-item FFQ was administered every 4 years to update diet. Using a similar FFQ, dietary data was collected every four years from the NHS II participants starting in 1991 and from the HPFS participants starting in 1986. In all FFQs, participants were asked how often (from “never or less than once per month” to “6 or more times per day”) on average they consumed a standard portion size of each food item during the previous year. The questionnaire items for coffee included “caffeinated coffee” and “decaffeinated coffee”. Consumption of total coffee was calculated as the sum of intakes of caffeinated and decaffeinated coffee. The validity and reproducibility of the FFQ has been described in detail elsewhere<sup>24-27</sup>. In brief, the validation study found a correlation coefficient of 0.78 between coffee intake assessed on the baseline FFQ and coffee intake assessed on four 1-week dietary records collected over a one year period<sup>26</sup>. As mean coffee consumption did not change in NHS 2 and decreased slightly in NHS and HPFS over time (Supplemental figure 1), we used baseline coffee consumption as primary exposure and further conducted several sensitivity analyses using updated dietary information.

### **Assessment of Covariates**

In the biennial follow-up questionnaires, updated information was collected on age, weight, smoking status, physical activity, medication use, family history of diabetes, and self-reported diagnosis of diseases, including hypertension, hypercholesterolemia, CVD, and cancer. For NHS and NHS 2 participants, we also ascertained data on menopausal status and postmenopausal hormone use. We calculated the Alternate Healthy Eating Index (aHEI) as an overall measure of diet quality using FFQ data<sup>28</sup>.

### **Assessment of Deaths**

Our primary end point was death from any cause. We performed systematic searches of the vital records of states and of the National Death Index. This search was supplemented by reports from family members and postal authorities. Using these methods, we were able to ascertain more than 98% of the deaths in each cohort<sup>29</sup>. A physician who was blinded to data on coffee consumption and other risk factors reviewed death certificates and medical records to classify the cause of death according to the eighth and ninth revisions of the *International Classification of Diseases*. Deaths were grouped into nine major categories (Supplemental Table 1).

### **Statistical Analysis**

We calculated each individual's person-time from the date of the return of the baseline questionnaire to the date of death or the end of follow-up (31 December 2012 for the NHS, 31 December 2012 for the NHS 2, and 31 December 2012 for the HPFS), whichever came first. We used Cox proportional hazards regression models to examine the association between coffee consumption (five categories) and risk of mortality. The regression models included calendar time in 2-y intervals as the time scale, and were stratified by age in years. In the multivariable

analysis, we further adjusted for body mass index (BMI), physical activity, overall dietary pattern (aHEI), total energy intake, smoking status, sugar-sweetened beverage consumption and alcohol consumption, all of which were updated from follow-up questionnaires. We additionally adjusted for baseline hypertension, hypercholesterolemia, and diabetes status in both men and women, and menopausal status, and postmenopausal hormone use among women.

We also used restricted cubic splines with 3 knots to flexibly model the association between coffee consumption and risk of mortality. To test for a potential non-linear association between coffee consumption and risk of mortality, a likelihood ratio test (LRT) was used comparing the model with only the linear term of coffee consumption to the model with both the linear and the cubic spline terms, with  $P$  value  $< 0.05$  denoting significant non-linearity. All analyses were performed separately in each cohort, and then pooled to obtain the overall hazard ratio using a fixed-effects model.

Stratified analyses were conducted according to BMI ( $\leq 25 \text{ kg/m}^2$ ,  $> 25 \text{ kg/m}^2$ ), age ( $\leq 70\text{y}$ ,  $> 70\text{y}$ ), aHEI ( $\leq$  median score,  $>$  median score), physical activity ( $\leq$  median,  $>$  median), smoking status (never smokers, ever smokers), sex (male, female), and individual cohort. We tested for potential effect modification by these stratification variables by including interaction terms between the exposure and potential effect modifier in the multivariate adjusted model, and conducting a likelihood ratio test (LRT) comparing the models with and without interaction terms.

The proportional hazard assumption of the Cox model was tested by adding interaction terms between exposure and the dichotomized indicator of time intervals to the multivariate adjusted model within each cohort, and conducting a likelihood ratio test comparing the models with and without interaction terms.

All statistical tests were 2-sided and performed using SAS version 9.2 for UNIX (SAS Institute Inc).

## RESULTS

### Coffee Consumption and Dietary and Lifestyle Factors

The percentages of never coffee drinkers were 12% in NHS, 30% in NHS 2, and 17% in HPFS. The percentages of those who drank more than 5 cups/d were 8% in NHS, 3% in NHS 2, and 5% in HPFS. There was a strong correlation between frequent coffee consumption and smoking status (Table 1). The proportions of never smokers among those who did not drink coffee were 63%, 80%, and 71% in NHS, NHS 2, and HPFS respectively, while the proportions of never smokers among those who drank more than 5 cups/d were 24%, 35%, and 25% in NHS, NHS 2, and HPFS. Those who drank coffee more frequently were also more likely to consume alcohol, and consumed less sugar-sweetened beverages and fruits, but more red meats.

### Coffee Consumption and All-cause Mortality

During 28 years of follow-up (1,894,292 person-years) among women in the NHS, we documented 17,468 deaths; during 21 years of follow up (1,882,464 person-years) among women in the NHS 2, we documented 2,056 deaths; during 26 years of follow-up (913,316 person-years) among men in the HPFS, we documented 12,432 deaths. In total, 31,956 deaths were recorded during 4,690,072 person-years of follow-up across all three cohorts.

Age-adjusted analysis showed that the highest categories of consumption of total and caffeinated coffee were associated with a higher risk of all-cause mortality across the three cohorts. The association between consumption of total, caffeinated, and decaffeinated coffee and all-cause

mortality attenuated significantly after further adjusting for smoking. Multivariate-adjusted analysis showed a non-linear association between consumption of total, caffeinated, and decaffeinated coffee and all-cause mortality ( $P$  values for non-linearity using LRT  $< 0.001$ ;  $P$  values for non-linear trend  $< 0.001$ ) (Table 2). Relative to no consumption of coffee, the pooled hazard ratio for death was 0.95 (95% CI: 0.91 to 0.99) for  $\leq 1$  cup of total coffee per day, 0.91 (95% CI: 0.88 to 0.95) for 1.1 - 3 cups per day, 0.93 (95% CI: 0.89 to 0.97) for 3.1 - 5 cups per day, and 1.02 (95% CI: 0.96 to 1.07) for  $> 5$  cups per day. Similar results were found when caffeinated and decaffeinated coffee was examined separately. Examining the three cohorts individually, the non-linear associations between consumption of total coffee, caffeinated coffee, and decaffeinated coffee and risk of all-cause mortality were most pronounced in NHS (Table 2, Supplemental figures 2 - 4).

As smoking is a strong confounder of the coffee-mortality relationship, we repeated the analysis among never smokers only. In this analysis, 10,505 deaths were documented during 2,451,970 person-years of follow-up after pooling data from the three cohorts. Overall, the association of total coffee, caffeinated coffee, and decaffeinated coffee consumption with risk of all-cause mortality changed from a non-linear association in the overall population to a linear inverse association when restricting to never smokers (total coffee:  $P$  for non-linearity = 0.32,  $P$  for linear trend  $< 0.001$ ; Caffeinated coffee:  $P$  for non-linearity = 0.40,  $P$  for linear trend  $< 0.001$ ; Decaffeinated coffee:  $P$  for non-linearity = 0.18,  $P$  for linear trend = 0.02) (Table 3, Figure 1).

#### Coffee Consumption and Cause-specific Mortality

The association between coffee consumption and leading causes of mortality was further investigated (Supplemental table 2). In the whole population, coffee consumption was inversely

associated with risk of mortality due to CVD, non-linearly associated with risk mortality due to type 2 diabetes, and positively associated with risks of mortality due to lung cancer and respiratory diseases (Supplemental table 3). However, when restricting to never smokers, coffee consumption was no longer associated with risk of mortality due to lung cancer and respiratory disease, but was inversely associated only with risks of mortality due to CVD, neurological disease, and suicide. No associations of coffee consumption with risks of mortality due to colorectal cancer and breast cancer were found (Table 4). The associations of coffee consumption with cause-specific mortality were similar to the associations shown in the whole population (Supplemental table 4).

In the total population, a 1-cup per day increment in coffee consumption was positively associated with mortality due to lung cancer ( $P < 0.0001$ ) and respiratory disease ( $P < 0.05$ ), and inversely associated with mortality due to CHD, stroke, neurological disease, and type 2 diabetes ( $P < 0.05$ ). However, after restricting to never smokers, the positive association disappeared for lung cancer and respiratory disease and significant inverse associations remained for mortality due to CHD, neurological disease, and suicide ( $P < 0.05$ ) (Figure 2).

### Stratified Analysis

Significant interactions were found between coffee consumption and risk of mortality by age ( $P$  for interaction = 0.003) and smoking ( $P$  for interaction = 0.015) (Supplemental table 5). The association appeared to be stronger among those aged  $< 70$  years than older individuals and was stronger among never smokers than smokers. There were no significant differences in the associations between coffee consumption and risk of total mortality when stratified by aHEI score, BMI, physical activity, sex, and cohort.

The proportional hazard assumption was violated in NHS (but not in NHS 2 or HPFS), with a stronger association between coffee consumption and mortality in earlier time intervals.

However, non-linear associations between coffee consumption and risk of mortality were found in both subgroups stratified by time interval in NHS (Supplemental table 6). We further assessed the proportional hazard assumption among never smokers in NHS, and the proportional hazard assumption was no longer violated (P for interaction = 0.60).

### Sensitivity Analysis

We further evaluated the association between coffee consumption and mortality using cumulatively updated coffee consumption and stopping updating of coffee consumption when intermediate diseases developed (hypertension, hypercholesterolemia, type 2 diabetes, cancer, and CVD); using time-varying coffee consumption with 4-year lag; using time-varying coffee consumption adjusting for hypercholesterolemia as a time-varying covariate; and using baseline coffee consumption excluding the hypertensive or hypercholesterolemia cases at baseline. The associations between consumption of total coffee, caffeinated coffee, and decaffeinated coffee and risk of mortality did not change substantially in these analyses (Supplemental table 7-8, Supplemental figure 5 - 9).

To further evaluate whether the change of the association from non-linear in the whole population to inverse linear among never smokers was due to the differences in the composition of total mortality between the overall population and never smokers, Cox models with inverse probability weighting were applied in the never smokers assessing the association between coffee consumption and risk of mortality and the results did not change substantially (Supplemental table 9).



## DISCUSSION

In this analysis of three large ongoing cohort studies, we observed a non-linear association between coffee consumption and risk of mortality in the overall population, with moderate coffee consumption being associated with lower mortality risk, and high coffee consumption not being associated with mortality risk. Given that this association became linear and inverse after restricting to never smokers, it is likely that the non-linear association observed in the total population was due to the residual confounding by smoking. This was further strengthened by the observation that the positive association between coffee consumption and death due to lung cancer and respiratory diseases in the overall population, for both of which smoking is an important risk factor, disappeared when restricting to never smokers. The inverse association between coffee consumption and risk of mortality did not change substantially when using a weighted Cox model among never smokers, excluding the possibility that the different associations in overall population and never smokers were due to the different composition of total mortality. For both caffeinated and decaffeinated coffee consumption, the non-linear associations in the total population and the inverse associations among the never smokers decreased the possibility that the non-linear association was due to the biological effect of caffeine.

Our results for the associations between coffee consumption and cause-specific mortality are consistent with the associations between coffee consumption and cause-specific diseases from previous studies. Numerous prospective cohort studies have shown coffee consumption to be associated with lower risk of type 2 diabetes<sup>1</sup>. There are several plausible biological mechanisms that could explain this observation. The chlorogenic acid, lignans, quinides, trigonelline, and magnesium in coffee reduce insulin resistance and systematic inflammation<sup>30, 31</sup>

<sup>32-34 35</sup>. Chlorogenic acid may have this putative effect by reducing glucose absorption in the intestine by competitively inhibiting glucose-6-phosphate translocase and reducing sodium-dependent glucose transport in the brush border membrane vesicles <sup>36</sup>; by reducing oxidative stress as a result of its antioxidant properties; and by reducing liver glucose output <sup>37</sup>. In our study, an inverse association between coffee drinking and risk of mortality due to CVD was observed. Given that diabetes and CVD share common disease pathways, the mechanism of inverse association between coffee consumption and risk of CVD mortality might be similar to that for diabetes mortality. Studies have also shown coffee consumption to be associated with a lower risk of Parkinson's diseases (PD) <sup>6, 38, 39</sup>, which is consistent with our finding of an inverse association between coffee consumption and risk of neurological mortality. In a 1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine (MPTP) neurotoxin model of PD, caffeine was shown to attenuate MPTP - induced striatal dopamine loss, striatal dopamine transporter binding sites loss, and dopaminergic neurons loss, which might be mediated through A<sub>2A</sub> adenosine receptors <sup>40</sup>. Three published cohort studies have shown an inverse association between coffee consumption and risk of suicide <sup>17, 41, 42</sup>, however, one study showed a J-shaped association where heavy coffee consumption was associated with a higher risk of suicide <sup>43</sup>. Our study had shown an inverse association of both caffeinated and decaffeinated coffee consumption with risk of suicide in both the whole population and never smokers, indicating that coffee consumption might have antidepressant effects. Studies have shown an inverse association between coffee consumption and risk of liver diseases or risk of mortality due to liver diseases<sup>2, 44-47</sup>, however, no association of coffee consumption with risk of mortality due to liver diseases was found in our study, which might be due to the limited power given the small number of cases. Previous cohort studies

showed no association between coffee consumption and risk of colorectal cancer<sup>48, 49</sup>, which was consistent with our results.

Our results showed similar associations of caffeinated and decaffeinated coffee consumption with risk of total and cause-specific mortality in both the overall population and never smokers, further showing that other components in coffee besides caffeine might play a beneficial role mediating the association between long-term coffee consumption and risk of mortality. However, short-term metabolic studies have shown that caffeine could acutely increase blood pressure by antagonizing the adenosine A1 and A2A receptor<sup>50-52</sup>, and could also acutely adversely affect arterial stiffness and endothelium dependent vasodilation<sup>53, 54</sup>. Case crossover studies showed that coffee consumption transiently increased the risk of nonfatal myocardial infarction, ischemic stroke onset, and sudden cardiac death<sup>55-57</sup>. One cohort study assessed the association of coffee consumption with total mortality in subsequent 2 years among CVD participants, and no association was found<sup>14</sup>. However, it is still difficult to differentiate acute effects from long-term effects of habitual coffee consumption.

Our analysis has several strengths. The large sample size, long follow-up time, and a large number of deaths provided sufficient power to detect a non-linear association in the overall population and to perform further analyses among never smokers. The large number of deaths also allowed us to conduct analyses on cause-specific mortality. In addition, we had detailed measures of both caffeinated and decaffeinated coffee consumption as well as other dietary and lifestyle factors.

Several potential limitations also need to be considered. First, given the observational nature of the study design, we could not directly establish a cause-effect relationship between coffee and

mortality. Second, assessment of coffee intake was based on FFQs and thus measurement errors are inevitable. However, our validation studies have demonstrated high validity (Pearson correlation = 0.74) of the coffee intake by the FFQs as compared to multiple week diet records, and high reproducibility (Pearson correlation = 0.80) by comparing two consecutive FFQs <sup>26</sup>. Moreover, coffee intake was also one of the food items showing the highest validity and reproducibility by the FFQs in Europe <sup>58</sup> and Asia <sup>59, 60</sup>, indicating that coffee was a beverage less prone to misreporting. Finally, since our cohort participants comprise medical and health professionals and the vast majority of them are white, the results may not be generalizable to other populations.

In conclusion, regular consumption of coffee was inversely associated with risk of total mortality and mortality due to CVD, and neurological disease. Similar associations of caffeinated and decaffeinated coffee consumption with risk of total and cause-specific mortality were found. Results from this and previous studies indicate that coffee consumption can be incorporated into a healthy lifestyle.

**Funding** The study was supported by research grants UM1 CA186107, UM1 CA176726, UM1 CA167552, P01 CA87969, P01 CA055075, R01 HL034594, HL088521, HL35464 and HL60712 from the National Institutes of Health.

### **Competing Interests**

None of the authors had any financial or personal conflict of interest to disclose. RVD received research funding from Nestec S.A., Vevey, Switzerland.

### **Acknowledgement**

We would like to thank the participants and staff of the NHS, NHS2, and HPFS for their valuable contributions as well as the following state cancer registries for their help: AL, AZ, AR, CA, CO, CT, DE, FL, GA, ID, IL, IN, IA, KY, LA, ME, MD, MA, MI, NE, NH, NJ, NY, NC, ND, OH, OK, OR, PA, RI, SC, TN, TX, VA, WA, WY. The authors assume full responsibility for analyses and interpretation of these data.

### **Reference**

1. Ding M, Bhupathiraju SN, Chen M, van Dam RM, Hu FB. Caffeinated and decaffeinated coffee consumption and risk of type 2 diabetes: A systematic review and a dose-response meta-analysis. *Diabetes Care*. 2014;37:569-586
2. Kurozawa Y, Ogimoto I, Shibata A, Nose T, Yoshimura T, Suzuki H, Sakata R, Fujita Y, Ichikawa S, Iwai N, Tamakoshi A, Group JS. Coffee and risk of death from hepatocellular carcinoma in a large cohort study in japan. *Br J Cancer*. 2005;93:607-610
3. Je Y, Giovannucci E. Coffee consumption and risk of endometrial cancer: Findings from a large up-to-date meta-analysis. *Int J Cancer*. 2012;131:1700-1710
4. Wilson KM, Kasperzyk JL, Rider JR, Kenfield S, van Dam RM, Stampfer MJ, Giovannucci E, Mucci LA. Coffee consumption and prostate cancer risk and progression in the health professionals follow-up study. *J Natl Cancer Inst*. 2011;103:876-884
5. Song F, Qureshi AA, Han J. Increased caffeine intake is associated with reduced risk of basal cell carcinoma of the skin. *Cancer Res*. 2012;72:3282-3289
6. Ross GW, Abbott RD, Petrovitch H, Morens DM, Grandinetti A, Tung KH, Tanner CM, Masaki KH, Blanchette PL, Curb JD, Popper JS, White LR. Association of coffee and caffeine intake with the risk of parkinson disease. *JAMA*. 2000;283:2674-2679
7. Ding M, Bhupathiraju SN, Satija A, van Dam RM, Hu FB. Long-term coffee consumption and risk of cardiovascular disease: A systematic review and a dose-response meta-analysis of prospective cohort studies. *Circulation*. 2014;129:643-659
8. Sugiyama K, Kuriyama S, Akhter M, Kakizaki M, Nakaya N, Ohmori-Matsuda K, Shimazu T, Nagai M, Sugawara Y, Hozawa A, Fukao A, Tsuji I. Coffee consumption and mortality due to all causes, cardiovascular disease, and cancer in japanese women. *J Nutr*. 2010;140:1007-1013
9. Gardener H, Rundek T, Wright CB, Elkind MS, Sacco RL. Coffee and tea consumption are inversely associated with mortality in a multiethnic urban population. *J Nutr*. 2013;143:1299-1308

10. Freedman ND, Park Y, Abnet CC, Hollenbeck AR, Sinha R. Association of coffee drinking with total and cause-specific mortality. *N Engl J Med*. 2012;366:1891-1904
11. Lopez-Garcia E, van Dam RM, Li TY, Rodriguez-Artalejo F, Hu FB. The relationship of coffee consumption with mortality. *Ann Intern Med*. 2008;148:904-914
12. Zhang W, Lopez-Garcia E, Li TY, Hu FB, van Dam RM. Coffee consumption and risk of cardiovascular diseases and all-cause mortality among men with type 2 diabetes. *Diabetes Care*. 2009;32:1043-1045
13. Zhang WL, Lopez-Garcia E, Li TY, Hu FB, van Dam RM. Coffee consumption and risk of cardiovascular events and all-cause mortality among women with type 2 diabetes. *Diabetologia*. 2009;52:810-817
14. Lopez-Garcia E, Rodriguez-Artalejo F, Li TY, Mukamal KJ, Hu FB, van Dam RM. Coffee consumption and mortality in women with cardiovascular disease. *Am J Clin Nutr*. 2011;94:218-224
15. Malerba S, Turati F, Galeone C, Pelucchi C, Verga F, La Vecchia C, Tavani A. A meta-analysis of prospective studies of coffee consumption and mortality for all causes, cancers and cardiovascular diseases. *Eur J Epidemiol*. 2013;28:527-539
16. Crippa A, Discacciati A, Larsson SC, Wolk A, Orsini N. Coffee consumption and mortality from all causes, cardiovascular disease, and cancer: A dose-response meta-analysis. *Am J Epidemiol*. 2014;180:763-775
17. Klatsky AL, Armstrong MA, Friedman GD. Coffee, tea, and mortality. *Ann Epidemiol*. 1993;3:375-381
18. LeGrady D, Dyer AR, Shekelle RB, Stamler J, Liu K, Paul O, Lepper M, Shryock AM. Coffee consumption and mortality in the Chicago Western Electric Company study. *Am J Epidemiol*. 1987;126:803-812
19. Lindsted KD, Kuzma JW, Anderson JL. Coffee consumption and cause-specific mortality. Association with age at death and compression of mortality. *J Clin Epidemiol*. 1992;45:733-742
20. Liu J, Sui X, Lavie CJ, Hebert JR, Earnest CP, Zhang J, Blair SN. Association of coffee consumption with all-cause and cardiovascular disease mortality. *Mayo Clin Proc*. 2013;88:1066-1074
21. Je Y, Giovannucci E. Coffee consumption and total mortality: A meta-analysis of twenty prospective cohort studies. *Br J Nutr*. 2014;111:1162-1173
22. O'Keefe JH, Bhatti SK, Patil HR, DiNicolantonio JJ, Lucan SC, Lavie CJ. Effects of habitual coffee consumption on cardiometabolic disease, cardiovascular health, and all-cause mortality. *J Am Coll Cardiol*. 2013;62:1043-1051
23. Crippa A, Discacciati A, Larsson SC, Wolk A, Orsini N. Coffee consumption and mortality from all causes, cardiovascular disease, and cancer: A dose-response meta-analysis. *Am J Epidemiol*. 2014
24. Feskanich D, Rimm EB, Giovannucci EL, Colditz GA, Stampfer MJ, Litin LB, Willett WC. Reproducibility and validity of food intake measurements from a semiquantitative food frequency questionnaire. *J Am Diet Assoc*. 1993;93:790-796
25. Rimm EB, Giovannucci EL, Stampfer MJ, Colditz GA, Litin LB, Willett WC. Reproducibility and validity of an expanded self-administered semiquantitative food frequency questionnaire among male health professionals. *Am J Epidemiol*. 1992;135:1114-1126; discussion 1127-1136
26. Salvini S, Hunter DJ, Sampson L, Stampfer MJ, Colditz GA, Rosner B, Willett WC. Food-based validation of a dietary questionnaire: The effects of week-to-week variation in food consumption. *Int J Epidemiol*. 1989;18:858-867
27. Willett WC, Sampson L, Stampfer MJ, Rosner B, Bain C, Witschi J, Hennekens CH, Speizer FE. Reproducibility and validity of a semiquantitative food frequency questionnaire. *Am J Epidemiol*. 1985;122:51-65

28. McCullough ML, Willett WC. Evaluating adherence to recommended diets in adults: The alternate healthy eating index. *Public Health Nutr.* 2006;9:152-157
29. Rich-Edwards JW, Corsano KA, Stampfer MJ. Test of the national death index and equifax nationwide death search. *Am J Epidemiol.* 1994;140:1016-1019
30. van Dam RM, Hu FB. Coffee consumption and risk of type 2 diabetes: A systematic review. *JAMA.* 2005;294:97-104
31. Williams CJ, Fargnoli JL, Hwang JJ, Van Dam RM, Blackburn GL, Hu FB, Mantzoros CS. Coffee consumption is associated with higher plasma adiponectin concentrations in women with or without type 2 diabetes: A prospective cohort study. *Diabetes Care.* 2008;31:504-507
32. van Dam RM. Coffee and type 2 diabetes: From beans to beta-cells. *Nutr Metab Cardiovasc Dis.* 2006;16:69-77
33. Van Dijk AE, Olthof MR, Meeuse JC, Seebus E, Heine RJ, Van Dam RM. Acute effects of decaffeinated coffee and the major coffee components chlorogenic acid and trigonelline on glucose tolerance. *Diabetes Care.* 2009;32:1023-1025
34. Greenberg JA, Boozer CN, Geliebter A. Coffee, diabetes, and weight control. *Am J Clin Nutr.* 2006;84:682-693
35. Lopez-Ridaura R, Willett WC, Rimm EB, Liu S, Stampfer MJ, Manson JE, Hu FB. Magnesium intake and risk of type 2 diabetes in men and women. *Diabetes Care.* 2004;27:134-140
36. Arion WJ, Canfield WK, Ramos FC, Schindler PW, Burger HJ, Hemmerle H, Schubert G, Below P, Herling AW. Chlorogenic acid and hydroxynitrobenzaldehyde: New inhibitors of hepatic glucose 6-phosphatase. *Arch Biochem Biophys.* 1997;339:315-322
37. Svilaas A, Sakhi AK, Andersen LF, Svilaas T, Strom EC, Jacobs DR, Jr., Ose L, Blomhoff R. Intakes of antioxidants in coffee, wine, and vegetables are correlated with plasma carotenoids in humans. *J Nutr.* 2004;134:562-567
38. Saaksjarvi K, Knekt P, Rissanen H, Laaksonen MA, Reunanen A, Mannisto S. Prospective study of coffee consumption and risk of parkinson's disease. *Eur J Clin Nutr.* 2008;62:908-915
39. Ascherio A, Zhang SM, Hernan MA, Kawachi I, Colditz GA, Speizer FE, Willett WC. Prospective study of caffeine consumption and risk of parkinson's disease in men and women. *Ann Neurol.* 2001;50:56-63
40. Chen JF, Xu K, Petzer JP, Staal R, Xu YH, Beilstein M, Sonsalla PK, Castagnoli K, Castagnoli N, Jr., Schwarzschild MA. Neuroprotection by caffeine and a(2a) adenosine receptor inactivation in a model of parkinson's disease. *J Neurosci.* 2001;21:RC143
41. Lucas M, O'Reilly EJ, Pan A, Mirzaei F, Willett WC, Okereke OI, Ascherio A. Coffee, caffeine, and risk of completed suicide: Results from three prospective cohorts of american adults. *World J Biol Psychiatry.* 2014;15:377-386
42. Kawachi I, Willett WC, Colditz GA, Stampfer MJ, Speizer FE. A prospective study of coffee drinking and suicide in women. *Arch Intern Med.* 1996;156:521-525
43. Tanskanen A, Tuomilehto J, Viinamaki H, Vartiainen E, Lehtonen J, Puska P. Heavy coffee drinking and the risk of suicide. *Eur J Epidemiol.* 2000;16:789-791
44. Larsson SC, Wolk A. Coffee consumption and risk of liver cancer: A meta-analysis. *Gastroenterology.* 2007;132:1740-1745
45. Bravi F, Bosetti C, Tavani A, Bagnardi V, Gallus S, Negri E, Franceschi S, La Vecchia C. Coffee drinking and hepatocellular carcinoma risk: A meta-analysis. *Hepatology.* 2007;46:430-435
46. Lai GY, Weinstein SJ, Albanes D, Taylor PR, McGlynn KA, Virtamo J, Sinha R, Freedman ND. The association of coffee intake with liver cancer incidence and chronic liver disease mortality in male smokers. *Br J Cancer.* 2013;109:1344-1351
47. Tverdal A, Skurtveit S. Coffee intake and mortality from liver cirrhosis. *Ann Epidemiol.* 2003;13:419-423

48. Tavani A, La Vecchia C. Coffee, decaffeinated coffee, tea and cancer of the colon and rectum: A review of epidemiological studies, 1990-2003. *Cancer Causes Control*. 2004;15:743-757
49. Je Y, Liu W, Giovannucci E. Coffee consumption and risk of colorectal cancer: A systematic review and meta-analysis of prospective cohort studies. *Int J Cancer*. 2009;124:1662-1668
50. Nurminen ML, Niittynen L, Korpela R, Vapaatalo H. Coffee, caffeine and blood pressure: A critical review. *Eur J Clin Nutr*. 1999;53:831-839
51. Mesas AE, Leon-Munoz LM, Rodriguez-Artalejo F, Lopez-Garcia E. The effect of coffee on blood pressure and cardiovascular disease in hypertensive individuals: A systematic review and meta-analysis. *Am J Clin Nutr*. 2011;94:1113-1126
52. Fredholm BB, Battig K, Holmen J, Nehlig A, Zvartau EE. Actions of caffeine in the brain with special reference to factors that contribute to its widespread use. *Pharmacol Rev*. 1999;51:83-133
53. Karatzis E, Papaioannou TG, Aznaouridis K, Karatzi K, Stamatelopoulos K, Zampelas A, Papamichael C, Lekakis J, Mavrikakis M. Acute effects of caffeine on blood pressure and wave reflections in healthy subjects: Should we consider monitoring central blood pressure? *Int J Cardiol*. 2005;98:425-430
54. Papamichael CM, Aznaouridis KA, Karatzis EN, Karatzi KN, Stamatelopoulos KS, Vamvakou G, Lekakis JP, Mavrikakis ME. Effect of coffee on endothelial function in healthy subjects: The role of caffeine. *Clin Sci (Lond)*. 2005;109:55-60
55. Baylin A, Hernandez-Diaz S, Kabagambe EK, Siles X, Campos H. Transient exposure to coffee as a trigger of a first nonfatal myocardial infarction. *Epidemiology*. 2006;17:506-511
56. Selb Semerl J, Selb K. Coffee and alcohol consumption as triggering factors for sudden cardiac death: Case-crossover study. *Croat Med J*. 2004;45:775-780
57. Mostofsky E, Schlaug G, Mukamal KJ, Rosamond WD, Mittleman MA. Coffee and acute ischemic stroke onset: The stroke onset study. *Neurology*. 2010;75:1583-1588
58. Date C, Fukui M, Yamamoto A, Wakai K, Ozeki A, Motohashi Y, Adachi C, Okamoto N, Kurosawa M, Tokudome Y, Kurisu Y, Watanabe Y, Ozasa K, Nakagawa S, Tokui N, Yoshimura T, Tamakoshi A. Reproducibility and validity of a self-administered food frequency questionnaire used in the jacc study. *J Epidemiol*. 2005;15 Suppl 1:S9-23
59. Esfahani FH, Asghari G, Mirmiran P, Azizi F. Reproducibility and relative validity of food group intake in a food frequency questionnaire developed for the tehran lipid and glucose study. *J Epidemiol*. 2010;20:150-158
60. Ferraroni M, Tavani A, Decarli A, Franceschi S, Parpinel M, Negri E, La Vecchia C. Reproducibility and validity of coffee and tea consumption in Italy. *Eur J Clin Nutr*. 2004;58:674-680



**Table 1. Age-adjusted baseline characteristics of participants by frequency of total coffee consumption (including caffeinated and decaffeinated coffee) in NHS, NHS 2, and HPFS**

	NHS (1984)					NHS 2 (1991)					HPFS (1986)				
Cups/d	0	≤ 1	1.1 - 3	3.1-5	> 5	0	≤ 1	1.1 - 3	3.1-5	> 5	0	≤ 1	1.1 - 3	3.1-5	> 5
N	9,233	14,740	30,420	14,760	5,737	27,888	22,837	29,239	10,049	3,041	6,863	11,402	14,264	5,861	2,167
Age (year)	48.4	50.5	50.7	50.5	50.0	35.2	35.5	37.6	38.0	38	50.9	54.0	53.7	52.9	52.0
Caffeinated coffee (cups/d)	0	0.4	1.7	3.2	4.7	0	0.4	1.9	3.4	5.15	0	0.4	1.6	3.0	4.3
Decaffeinated coffee (cups/d)	0	0.3	0.6	1.2	1.6	0	0.2	0.4	0.9	1.1	0	0.3	0.7	1.3	2.0
Physical activity (MET-h/wk)	14.1	14.1	14.4	13.8	13.3	23.1	24.8	25.5	24.2	25.6	21.7	22.2	21.1	20.3	18.2
aHEI <sup>#</sup>	46.9	48.0	47.8	47.8	47.3	46.0	49.0	50.2	49.7	48.5	51.9	53.7	52.7	51.7	50.2
Total energy intake (kcal/d)	1720	1684	1748	1786	1815	1779	1768	1790	1836	1883	1934	1887	1973	2026	2086
Sugar-sweetened beverages (serving/d)	1.30	1.2	1.1	1.0	0.9	1.4	1.2	1.1	0.9	0.9	1.4	1.3	1.1	1.0	0.9
Alcohol (grams/d)	3.7	5.8	7.8	7.7	7.1	1.6	2.7	4.3	4.4	4.2	5.7	10.0	13.1	14.1	14.8
Dairy (serving/d)	1.9	1.9	2.0	2.07	2.1	2.1	2.2	2.4	2.4	2.6	2.2	2.1	2.2	2.4	2.5
Fruits (serving/d)	2.2	2.3	2.2	2.1	1.9	1.8	2.0	1.9	1.8	1.6	2.6	2.4	2.2	2.1	2.0
Vegetables (serving/d)	3.0	3.0	3.1	3.1	3.1	3.2	3.5	3.6	3.7	3.8	3.0	3.0	3.0	3.0	3.0
Meats (serving/d)	1.4	1.4	1.5	1.5	1.6	1.4	1.4	1.4	1.4	1.5	1.4	1.4	1.6	1.6	1.8
Fish (serving/d)	0.2	0.3	0.3	0.3	0.3	0.2	0.2	0.2	0.2	0.2	0.3	0.3	0.3	0.3	0.3

BMI (kg/m <sup>2</sup> )	24.4	24.2	23.8	23.6	23.4	25.4	24.7	24.2	24.4	24.8	24.7	24.7	25.0	25.2	25.
Never smokers, %	63	53	42	34	24	80	70	57	47	35	71	54	42	33	25
Hypertension, %	9	10	8	7	6	4	4	3	3	3	18	21	20	19	17
Hypercholesterolemia, %	4	4	4	3	3	10	10	9	9	11	10	11	11	10	11
Postmenopausal women, %	48	48	48	49	49	3	3	3	3	4	NA	NA	NA	NA	NA
Current postmenopausal hormone use, (% among total women)	15	14	14	13	12	3	3	3	3	3	NA	NA	NA	NA	NA

aHEI: alternative healthy eating index; BMI: body mass index.

# aHEI ranges from 0 – 100, with a higher score indicating healthier diet.

**Table 2. HRs (95% CI) for the association between consumption of total coffee, caffeinated coffee, and decaffeinated coffee and risk of mortality**

<b>Total coffee</b> Intake (cups/d)	0	≤ 1 cup/d	1.1-3 cups/d	3.1-5 cups/d	>5 cups/d	Per cup increase	P for non- linearity*	P for linear trend
Cases/Person- time <i>NHS</i>	4,166/ 958,267	7,826/ 1,086,683	12,198/ 1,681,922	5,456/ 709,646	2,310/ 253,554			
Age-adjusted	1.00	0.98 (0.93, 1.04)	0.91 (0.87, 0.96)	0.96 (0.91, 1.02)	1.24 (1.16, 1.33)	1.02 (1.01, 1.03)	< 0.001	< 0.001
Age and smoking-adjusted	1.00	0.94 (0.89, 0.99)	0.82 (0.78, 0.86)	0.81 (0.76, 0.86)	0.92 (0.86, 0.98)	0.98 (0.97, 0.99)	< 0.001	< 0.001
Multivariate- adjusted model <i>NHS 2</i>	1.00	0.94 (0.88, 0.99)	0.90 (0.86, 0.95)	0.93 (0.88, 0.99)	1.02 (0.95, 1.09)	1.00 (0.99, 1.01)	< 0.001	0.50
Age-adjusted	1.00	0.89 (0.79, 1.00)	0.83 (0.74, 0.92)	0.93 (0.80, 1.07)	1.37 (1.12, 1.67)	0.98 (0.95, 1.01)	0.003	0.23
Age and smoking-adjusted	1.00	0.86 (0.76, 0.97)	0.74 (0.66, 0.83)	0.74 (0.64, 0.86)	0.92 (0.75, 1.13)	0.93 (0.90, 0.96)	0.015	< 0.001
Multivariate- adjusted model <i>HPFS</i>	1.00	0.91 (0.81, 1.03)	0.84 (0.75, 0.95)	0.86 (0.74, 1.01)	1.02 (0.83, 1.26)	0.96 (0.93, 1.00)	0.17	0.03
Age-adjusted	1.00	1.06 (1.00, 1.12)	1.05 (0.99, 1.11)	1.09 (1.02, 1.16)	1.30 (1.19, 1.42)	1.02 (1.00, 1.03)	0.98	0.007
Age and smoking-adjusted	1.00	1.00 (0.94, 1.06)	0.95 (0.90, 1.01)	0.94 (0.88, 1.01)	1.05 (0.96, 1.15)	0.98 (0.94, 1.02)	0.24	0.36
Multivariate- adjusted model <i>Pooled</i>	1.00	1.00 (0.94, 1.06)	0.97 (0.90, 1.01)	0.95 (0.88, 1.02)	1.02 (0.93, 1.12)	0.99 (0.97, 1.00)	0.13	0.04
Multivariate- adjusted model	1.00	0.95 (0.91, 0.99)	0.91 (0.88, 0.95)	0.93 (0.89, 0.97)	1.02 (0.96, 1.07)	0.98 (0.97, 0.99)	< 0.001	< 0.001

<b>Caffeinated coffee</b>								
Intake (cups/d)	0	≤ 1 cup/d	1.1-3 cups/d	3.1-5 cups/d	>5 cups/d	Per cup increase	P for non-linearity*	P for linear trend
Cases/Person-time <i>NHS</i>	8,615/ 1,454,869	10,105/ 1,404,192	8,495/ 1,255,722	3,304/ 419,049	1,437/ 156,238			
Multivariate-adjusted model <i>NHS 2</i>	1.00	0.96 (0.92, 1.00)	0.92 (0.89, 0.96)	1.01 (0.96, 1.07)	1.07 (0.99, 1.14)	1.00 (0.99, 1.01)	< 0.001	0.78
Multivariate-adjusted model <i>HPFS</i>	1.00	0.97 (0.86, 1.09)	0.91 (0.80, 1.02)	0.99 (0.83, 1.17)	1.11 (0.88, 1.40)	0.99 (0.96, 1.02)	0.38	0.47
Multivariate-adjusted model <i>Pooled</i>	1.00	1.00 (0.96, 1.05)	0.94 (0.90, 0.99)	1.00 (0.93, 1.07)	1.11 (1.00, 1.24)	0.99 (0.98, 1.00)	0.046	0.11
Multivariate-adjusted model	1.00	0.97 (0.94, 1.00)	0.93 (0.90, 0.96)	1.00 (0.96, 1.05)	1.08 (1.02, 1.14)	0.98 (0.97, 0.99)	0.015	< 0.001
<b>Decaffeinated coffee</b>								
Intake (mg/d)	0	≤ 1 cup/d	1.1-3 cups/d	>3 cups/d		Per cup increase	P for non-linearity*	P for linear trend
Cases/Person-time <i>NHS</i>	16,393/ 2,607,891	10,637/ 1,516,930	3,777/ 445,908	1,149/ 119,341				
Multivariate-adjusted model <i>NHS 2</i>	1.00	0.94 (0.90, 0.97)	0.92 (0.87, 0.96)	0.96 (0.89, 1.04)		0.96 (0.95, 0.98)	0.008	< 0.001
Multivariate-adjusted model <i>HPFS</i>	1.00	0.83 (0.74, 0.92)	0.86 (0.70, 1.04)	0.93 (0.64, 1.35)		0.92 (0.86, 1.00)	0.009	0.035
Multivariate-adjusted model	1.00	0.91 (0.88, 0.95)	0.92 (0.87, 0.98)	0.91 (0.83, 1.01)		0.97 (0.95, 0.99)	0.006	0.014

<i>Pooled</i>							
Multivariate-adjusted model	1.00	0.92 (0.89, 0.94)	0.91 (0.88, 0.94)	0.94 (0.88, 1.00)	0.96 (0.94, 0.97)	< 0.001	< 0.001

\* A likelihood ratio test was performed.

Multivariate-adjusted model: further adjusted for baseline disease status (hypertension, hypercholesterolemia, diabetes), BMI (< 20.9, 21-22.9, 23-24.9, 25-29.9, 30-34.9,  $\geq 35$  kg/m<sup>2</sup>), physical activity (< 3, 3-8.9, 9-17.9, 18-26.9,  $\geq 27$  MET-h/wk), overall dietary pattern (aHEI score, in quintiles), total energy intake (quintiles), smoking status (never, former (1 - 4 cigarettes/d), former (5 - 14 cigarettes/d), former (15 - 24 cigarettes/d), former (25 - 34 cigarettes/d), former (35 - 44 cigarettes/d), former ( $\geq 45$  cigarettes/d), former (unknown cigarettes/d), current (1 - 4 cigarettes/d), current (5 - 14 cigarettes/d), current (15 - 24 cigarettes/d), current (25 - 34 cigarettes/d), current (35 - 44 cigarettes/d), current ( $\geq 45$  cigarettes/d), current (unknown cigarettes/d)), sugar-sweetened beverage consumption (quintiles) and alcohol consumption (0, 0-5, 5-10, 10-15,  $\geq 15$  g/d). We additionally adjusted for menopausal status (yes vs. no), and postmenopausal hormone use (yes vs. no) for women. Caffeinated and decaffeinated coffee adjusted for each other.

**Table 3. HRs (95% CI) for the association between consumption of total coffee, caffeinated coffee, and decaffeinated coffee and risk of mortality among never smokers**

Total coffee								
Intake (cups/d)	0	≤ 1 cup/d	1.1-3 cups/d	3.1-5 cups/d	>5 cups/d	Per cup increase	P for non-linearity*	P for linear trend
Cases/Person-time	2,190/ 704,010	3,032/ 638,687	3,759/ 779,966	1,211/ 262,069	313/ 67,238			
<i>NHS</i>								
Age-adjusted model	1.00	0.92 (0.85, 0.99)	0.80 (0.74, 0.86)	0.73 (0.66, 0.80)	0.76 (0.65, 0.88)	0.93 (0.91, 0.95)	0.09	< 0.001
Multivariate-adjusted model	1.00	0.92 (0.85, 0.99)	0.88 (0.81, 0.95)	0.84 (0.76, 0.92)	0.82 (0.71, 0.96)	0.96 (0.95, 0.98)	0.09	< 0.001
<i>NHS 2</i>								
Age-adjusted model	1.00	0.84 (0.71, 0.98)	0.84 (0.72, 0.98)	0.76 (0.60, 0.97)	0.85 (0.55, 1.32)	0.95 (0.91, 1.00)	0.20	0.07
Multivariate-adjusted model	1.00	0.89 (0.76, 1.05)	0.98 (0.84, 1.15)	0.88 (0.69, 1.13)	0.91 (0.59, 1.42)	1.00 (0.95, 1.05)	0.77	0.96
<i>HPFS</i>								
Age-adjusted model	1.00	0.99 (0.90, 1.08)	0.96 (0.88, 1.05)	0.83 (0.73, 0.95)	1.06 (0.84, 1.33)	0.96 (0.94, 0.99)	0.20	0.005
Multivariate-adjusted model	1.00	0.98 (0.89, 1.08)	0.97 (0.88, 1.06)	0.84 (0.73, 0.97)	0.98 (0.78, 1.24)	0.96 (0.94, 0.99)	0.38	0.009
<i>Pooled</i>								
Multivariate-adjusted model	1.00	0.94 (0.89, 0.99)	0.92 (0.87, 0.97)	0.85 (0.79, 0.92)	0.88 (0.78, 0.99)	0.97 (0.95, 0.98)	0.32	< 0.001
Caffeinated coffee								
Intake (cups/d)	0	≤ 1 cup/d	1.1-3 cups/d	3.1-5 cups/d	>5 cups/d	Per cup increase	P for non-linearity*	P for linear trend
Cases/Person-time	3,702/ 962,732	3,637/ 771,531	2,409/ 542,629	595/ 138,375	162/ 36,700			
<i>NHS</i>								

Multivariate-adjusted model <i>NHS 2</i>	1.00	0.95 (0.89, 1.01)	0.89 (0.83, 0.96)	0.84 (0.75, 0.94)	0.82 (0.67, 1.00)	0.96 (0.94, 0.98)	0.34	< 0.001
Multivariate-adjusted model <i>HPFS</i>	1.00	0.99 (0.85, 1.15)	1.07 (0.90, 1.27)	1.05 (0.78, 1.42)	1.04 (0.60, 1.81)	1.02 (0.98, 1.08)	0.95	0.42
Multivariate-adjusted model <i>Pooled</i>	1.00	1.03 (0.95, 1.11)	0.92 (0.84, 1.02)	0.95 (0.80, 1.14)	1.06 (0.78, 1.44)	0.97 (0.95, 1.00)	0.76	0.08
Multivariate-adjusted model <b>Decaffeinated coffee</b>	1.00	0.98 (0.93, 1.03)	0.92 (0.87, 0.97)	0.89 (0.81, 0.97)	0.90 (0.77, 1.06)	0.97 (0.95, 0.99)	0.40	< 0.001
Intake (mg/d)	0	≤ 1 cup/d	1.1-3 cups/d	>3 cups/d		Per cup increase	P for non-linearity*	P for linear trend
Cases/Person-time <i>NHS</i>	5,436/ 1,423,563	3,775/ 799,017	1,084/ 191,219	210/ 38,171				
Multivariate-adjusted model <i>NHS 2</i>	1.00	0.97 (0.91, 1.03)	0.92 (0.85, 1.00)	0.97 (0.82, 1.14)		0.97 (0.93, 1.00)	0.56	0.03
Multivariate-adjusted model <i>HPFS</i>	1.00	0.82 (0.70, 0.95)	0.87 (0.65, 1.16)	0.70 (0.33, 1.47)		0.93 (0.83, 1.04)	0.06	0.21
Multivariate-adjusted model <i>Pooled</i>	1.00	0.96 (0.88, 1.03)	0.94 (0.83, 1.06)	0.70 (0.52, 0.95)		0.98 (0.93, 1.02)	0.61	0.33
Multivariate-adjusted model	1.00	0.95 (0.91, 0.99)	0.93 (0.86, 0.99)	0.89 (0.77, 1.02)		0.97 (0.94, 0.99)	0.18	0.02

\* A likelihood ratio test was performed.

Multivariate-adjusted model: further adjusted for baseline disease status (hypertension, hypercholesterolemia, diabetes), BMI (< 20.9, 21-22.9, 23-24.9, 25-29.9, 30-34.9,  $\geq 35$  kg/m<sup>2</sup>), physical activity (< 3, 3-8.9, 9-17.9, 18-26.9,  $\geq 27$  MET-h/wk), overall dietary pattern (AHEI score, in quintiles), total energy intake (quintiles), sugar-sweetened beverages consumption (quintiles) and alcohol consumption (0, 0-5, 5-10, 10-15,  $\geq 15$  g/d). We additionally adjusted for menopausal status (yes vs. no), and postmenopausal hormone use (yes vs. no) for women. Caffeinated and decaffeinated coffee adjusted for each other.



**Table 4. Multivariate HRs (95% CI) for the association between consumption of total coffee and risk of cause-specific mortality among never smokers**

	0	≤ 1 cup/d	1.1-3 cups/d	3.1-5 cups/d	>5 cups/d	P for nonlinearity*	P for linear trend
<b>CVD mortality</b>							
<b>(2587 cases)</b>							
Multivariate-adjusted model	1.00	0.95 (0.85, 1.07)	0.94 (0.84, 1.05)	0.81 (0.70, 0.95)	0.91 (0.71, 1.17)	0.77	0.004
<i>CHD mortality</i>							
<i>(1815 cases)</i>							
Multivariate-adjusted model	1.00	0.90 (0.78, 1.03)	0.90 (0.79, 1.03)	0.81 (0.68, 0.98)	0.89 (0.66, 1.20)	0.90	0.01
<i>Stroke mortality</i>							
<i>(656 cases)</i>							
Multivariate-adjusted model	1.00	1.04 (0.82, 1.31)	1.01 (0.80, 1.27)	0.76 (0.56, 1.04)	0.93 (0.56, 1.54)	0.81	0.06
<b>Cancer mortality</b>							
<b>(3664 cases)</b>							
Multivariate-adjusted model	1.00	0.99 (0.90, 1.09)	0.99 (0.90, 1.08)	0.88 (0.77, 0.99)	0.84 (0.68, 1.03)	0.34	0.08
<i>Colorectal cancer mortality</i>							
<i>(380 cases)</i>							
Multivariate-adjusted model	1.00	0.88 (0.65, 1.19)	0.92 (0.69, 1.23)	0.92 (0.63, 1.33)	0.94 (0.51, 1.75)	0.78	0.91
<i>Lung cancer mortality</i>							
<i>(217 cases)</i>							
Multivariate-adjusted model	1.00	1.11 (0.73, 1.69)	1.14 (0.76, 1.71)	1.15 (0.70, 1.90)	0.89 (0.37, 2.12)	0.32	0.82
<i>Pancreatic cancer mortality</i>							
<i>(321 cases)</i>							
Multivariate-adjusted model	1.00	1.47 (1.04, 2.06)	1.10 (0.78, 1.55)	1.06 (0.69, 1.62)	0.41 (0.15, 1.14)	0.056	0.06

<i>Breast cancer mortality</i> (567 cases)							
Multivariate-adjusted model	1.00	0.96 (0.76, 1.22)	0.91 (0.72, 1.14)	0.79 (0.58, 1.07)	0.62 (0.36, 1.09)	0.84	0.16
<i>Premenopausal breast cancer mortality</i> (77 cases)							
Multivariate-adjusted model	NA						
<i>Postmenopausal breast cancer mortality</i> (490 cases)							
Multivariate-adjusted model	1.00	0.94 (0.72, 1.21)	0.83 (0.64, 1.07)	0.79 (0.57, 1.08)	0.67 (0.38, 1.18)	0.57	0.10
<i>Ovary cancer mortality</i> (250 cases)							
Multivariate-adjusted model	1.00	0.73 (0.50, 1.05)	0.85 (0.61, 1.19)	0.51 (0.31, 0.83)	0.63 (0.30, 1.33)	0.71	0.20
<i>Endometrial cancer mortality</i> (115 cases)							
Multivariate-adjusted model	1.00	0.99 (0.55, 1.80)	1.24 (0.72, 2.15)	0.90 (0.43, 1.85)	2.17 (0.94, 5.05)	0.26	0.94
<i>Prostate cancer mortality</i> (210 cases)							
Multivariate-adjusted model	1.00	0.75 (0.51, 1.10)	0.87 (0.59, 1.28)	0.74 (0.42, 1.29)	0.83 (0.30, 2.35)	0.42	0.48
<b>Respiratory disease mortality</b> <b>(385 cases)</b>							
Multivariate-adjusted model	1.00	0.91 (0.67, 1.23)	0.88 (0.66, 1.19)	0.94 (0.65, 1.36)	0.62 (0.30, 1.31)	0.31	0.95
<b>Neurological disease mortality</b> <b>(243 cases)</b>							
Multivariate-adjusted model	1.00	0.91 (0.62, 1.32)	0.80 (0.55, 1.15)	0.63 (0.39, 1.01)	0.79 (0.38, 1.62)	0.83	0.004

<b>Diabetes mortality</b>							
<b>(128 cases)</b>							
Multivariate-adjusted model	1.00	0.96 (0.59, 1.56)	0.67 (0.40, 1.11)	0.76 (0.38, 1.49)	0.76 (0.26, 2.20)	0.19	0.09
<b>Injury mortality</b>							
<b>(233 cases)</b>							
Multivariate-adjusted model	1.00	1.01 (0.70, 1.46)	0.90 (0.62, 1.31)	0.71 (0.42, 1.22)	1.28 (0.62, 2.65)	0.28	0.22
<b>Suicide mortality</b>							
<b>(134 cases)</b>							
Multivariate-adjusted model	1.00	1.36 (0.88, 2.10)	0.73 (0.45, 1.21)	0.64 (0.30, 1.35)	0.80 (0.24, 2.65)	0.58	0.02
<b>Other disease mortality</b>							
<b>(3108 cases)</b>							
Multivariate-adjusted model	1.00	0.86 (0.77, 0.95)	0.81 (0.73, 0.90)	0.81 (0.71, 0.93)	0.94 (0.76, 1.16)	0.011	0.008

\* A likelihood ratio test was performed.

The model adjusted for age, baseline disease status (hypertension, hypercholesterolemia, diabetes), BMI (< 20.9, 21-22.9, 23-24.9, 25-29.9, 30-34.9,  $\geq 35$  kg/m<sup>2</sup>), physical activity (< 3, 3-8.9, 9-17.9, 18-26.9,  $\geq 27$  MET-h/wk), overall dietary pattern (AHEI score, in quintiles), total energy intake (quintiles), sugar-sweetened beverages consumption (quintiles) and alcohol consumption (0, 0-5, 5-10, 10-15,  $\geq 15$  g/d). We additionally adjusted for menopausal status (yes vs. no), and postmenopausal hormone use (yes vs. no) for women. Caffeinated and decaffeinated coffee adjusted for each other.

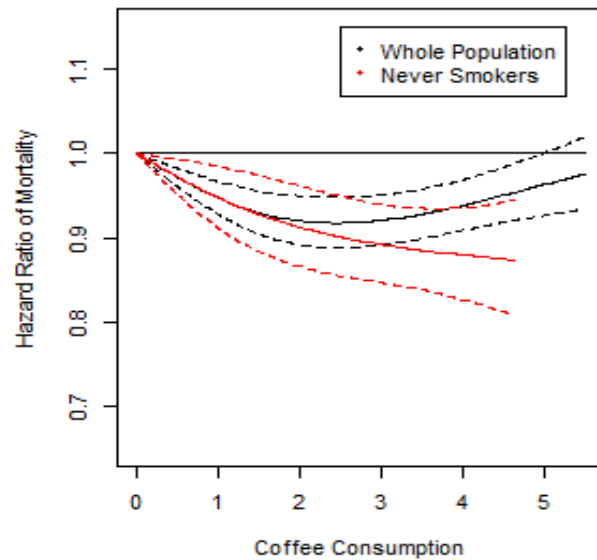


Figure 1a.

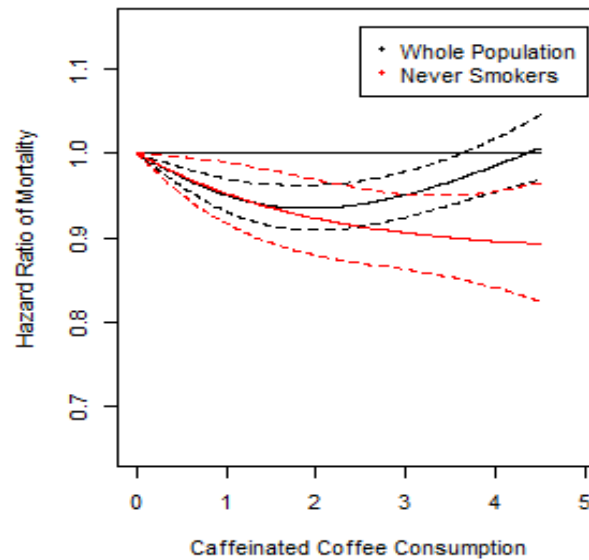


Figure 1b.

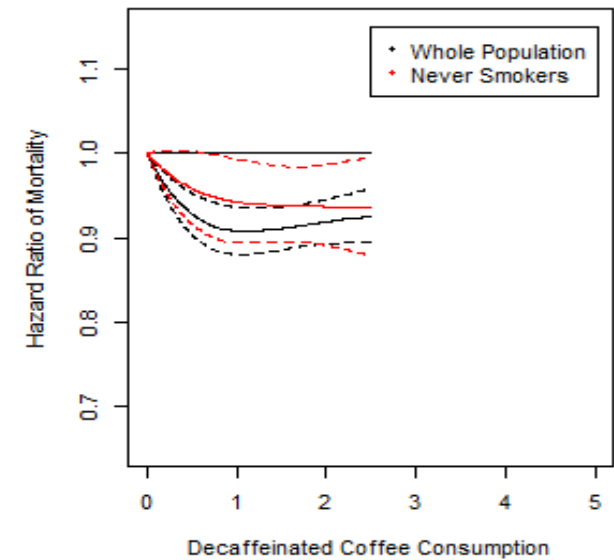


Figure 1c.

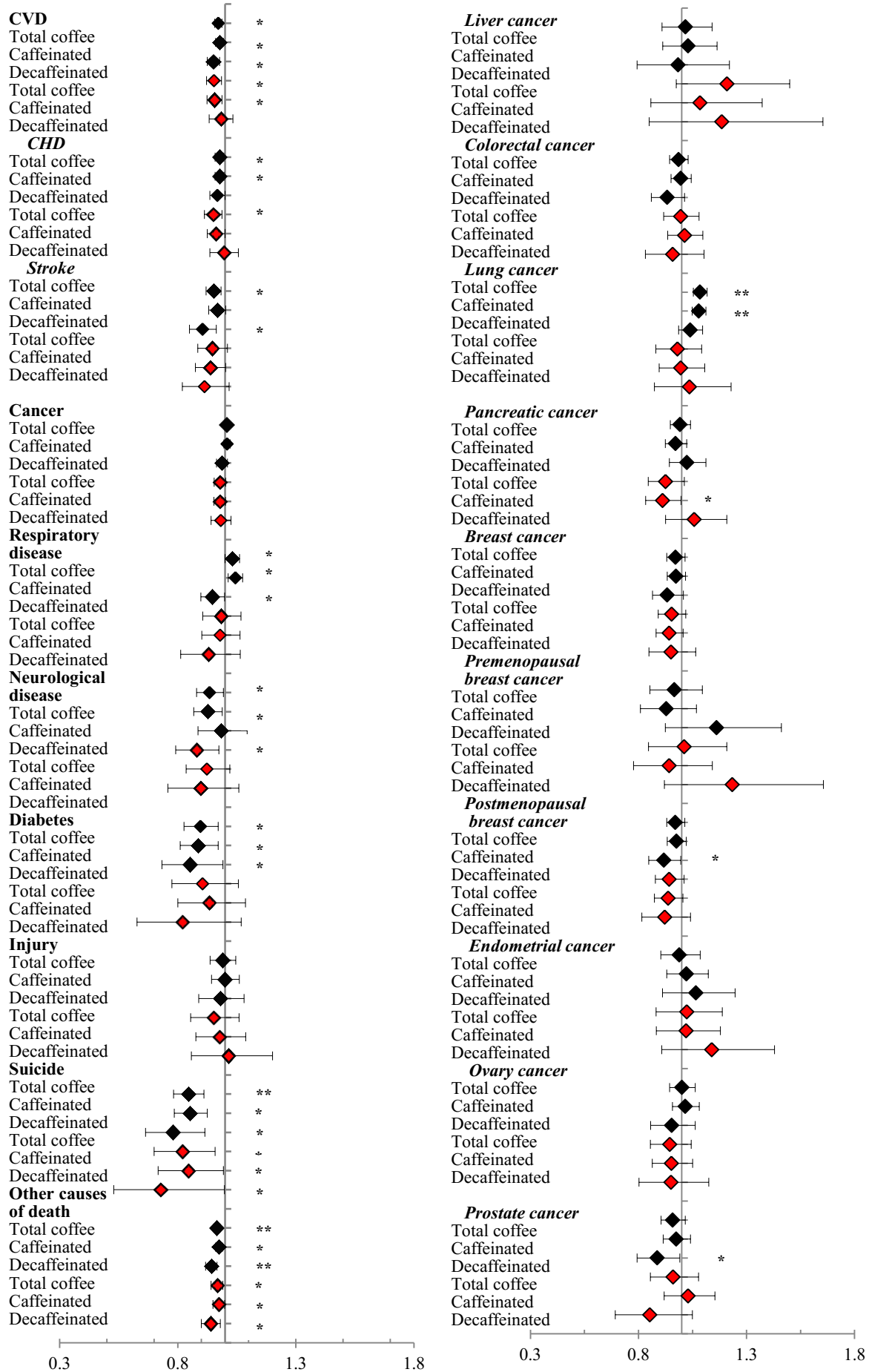
Figure 1. The association between coffee consumption and risk of mortality in the overall population and among never smokers pooled across the three cohorts. 1a. Total coffee consumption and risk of mortality 1b. Caffeinated coffee consumption and risk of mortality 1c. Decaffeinated coffee consumption and risk of mortality.

Multivariate-adjusted models adjusted for age, baseline disease status (hypertension, hypercholesterolemia, diabetes), BMI (< 20.9, 21-22.9, 23-24.9, 25-29.9, 30-34.9,  $\geq 35$  kg/m<sup>2</sup>), physical activity (< 3, 3-8.9, 9-17.9, 18-26.9,  $\geq 27$  MET-h/wk), smoking status (never, former (1 - 4 cigarettes/d), former (5 - 14 cigarettes/d), former (15 - 24 cigarettes/d), former (25 - 34 cigarettes/d), former (35 - 44 cigarettes/d), former ( $\geq 45$  cigarettes/d), former (unknown cigarettes/d), current (1 - 4 cigarettes/d), current (5 - 14 cigarettes/d), current (15 - 24 cigarettes/d), current (25 - 34 cigarettes/d), current (35 - 44 cigarettes/d), current ( $\geq 45$  cigarettes/d), current (unknown cigarettes/d)), overall

dietary pattern (AHEI score, in quintiles), total energy intake (quintiles), sugar-sweetened beverages consumption (quintiles) and alcohol consumption (0, 0-5, 5-10, 10-15,  $\geq$  15 g/d). We additionally adjusted for menopausal status (yes vs. no), and postmenopausal hormone use (yes vs. no) for women. Caffeinated and decaffeinated coffee adjusted for each other.

Figure 2. The association of a 1-cup per day increment in coffee consumption with risk of cause-specific mortality pooled across the three cohorts. The black squares stand for the overall population. The red squares stand for never smokers. \* P value < 0.05. \*\* P value < 0.001.

Multivariate-adjusted models adjusted for age, baseline disease status (hypertension, hypercholesterolemia, diabetes), BMI (< 20.9, 21-22.9, 23-24.9, 25-29.9, 30-34.9,  $\geq 35$  kg/m<sup>2</sup>), physical activity (< 3, 3-8.9, 9-17.9, 18-26.9,  $\geq 27$  MET-h/wk), smoking status (never, former (1 - 4 cigarettes/d), former (5 - 14 cigarettes/d), former (15 - 24 cigarettes/d), former (25 - 34 cigarettes/d), former (35 - 44 cigarettes/d), former ( $\geq 45$  cigarettes/d), former (unknown cigarettes/d), current (1 - 4 cigarettes/d), current (5 - 14 cigarettes/d), current (15 - 24 cigarettes/d), current (25 - 34 cigarettes/d), current (35 - 44 cigarettes/d), current ( $\geq 45$  cigarettes/d), current (unknown cigarettes/d)), overall dietary pattern (AHEI score, in quintiles), total energy intake (quintiles), sugar-sweetened beverages consumption (quintiles) and alcohol consumption (0, 0-5, 5-10, 10-15,  $\geq 15$  g/d). We additionally adjusted for menopausal status (yes vs. no), and postmenopausal hormone use (yes vs. no) for women. Caffeinated and decaffeinated coffee adjusted for each other.



**Supplemental Table 1. Categories for causes of death.**

Causes of death	ICD-8 code
Cardiovascular disease	390-458
Heart disease	390-429, 440-458
Stroke	430-438
Cancer	140-207
Colorectal cancer	153, 154
Lung cancer	162
Pancreatic cancer	157
Breast cancer	174
Ovary cancer	183
Prostate cancer	185
Respiratory disease	460-519
Diabetes	250
Neurological disease	290, 340, 342, 348
Injury	800-950, 959-999
Suicide	950-959
All other causes	The rest of the ICD codes



**Supplemental Table 2. The disease composition of overall mortality.**

	NHS	NHS 2	HPFS	Total population
Cardiovascular disease	3844	211	3781	7836
Coronary heart disease	2545	157	2966	5668
Stroke	1062	43	750	1855
Other diseases	4667	680	2925	8272
Respiratory disease	1383	52	924	2359
Injury	144	10	500	654
Neurological disease	460	48	22	530
Suicide	84	98	192	374
Type 2 diabetes	214	32	92	338
Renal disease	48	3	26	77
Cancer	6624	922	3970	11516
Lung cancer	1596	107	745	2448
Breast cancer	961	261	0	1222
Premenopausal breast cancer	42	83	0	125
Postmenopausal breast cancer	911	178	0	1089
Colorectal cancer	549	69	417	1035
Pancreatic cancer	487	49	338	874
Non-Hodgkin lymphoma	398	54	295	747
Prostate cancer	0	0	572	572
Ovary cancer	496	82	0	578
Brain cancer	195	58	167	420
Leukemia	193	30	175	398
Myeloma	162	7	145	314
Renal cell cancer	142	10	119	271
Bladder cancer	107	5	160	272
Skin cancer	105	26	126	257
Endometrial cancer	211	31	0	242
Esophagus cancer	71	5	133	209
Stomach cancer	102	9	87	198
Head and neck cancer	83	8	84	175

Liver cancer	72	5	60	137
Gall bladder cancer	89	13	32	134
Small intestine cancer	26	8	9	43
Cervix cancer	28	11	0	39
Hodgkin lymphoma	14	4	11	29
Total mortality	17,468	2,056	12,432	31,956

**Supplemental Table 3. HRs (95% CI) for the association between consumption of total coffee and risk of cause-specific mortality**

	0 cup/d	<1 cup/d	1-3 cups/d	3-5 cups/d	>5 cups/d	P <sub>non-linearity*</sub>	P for linear trend
<b>CVD mortality (7836 cases)</b>	1.00	1.01 (0.93, 1.09)	0.95 (0.87, 1.01)	0.88 (0.81, 0.96)	1.02 (0.91, 1.14)	0.49	< 0.001
<i>CHD mortality (5668 cases)</i>	1.00	0.98 (0.90, 1.07)	0.92 (0.85, 1.01)	0.90 (0.81, 0.99)	1.03 (0.91, 1.17)	0.21	0.01
<i>Stroke mortality (1855 cases)</i>	1.00	1.05 (0.89, 1.23)	0.95 (0.81, 1.11)	0.81 (0.67, 0.97)	1.00 (0.79, 1.25)	0.89	0.002
<b>Cancer mortality (11,516 cases)</b>	1.00	1.01 (0.94, 1.08)	1.01 (0.95, 1.08)	1.03 (0.95, 1.10)	1.16 (1.06, 1.27)	0.88	0.40
<i>Colorectal cancer mortality (1,035 cases)</i>	1.00	1.03 (0.83, 1.28)	0.95 (0.77, 1.17)	0.95 (0.75, 1.20)	0.94 (0.69, 1.29)	0.91	0.48
<i>Lung cancer mortality (2448 cases)</i>	1.00	1.03 (0.86, 1.23)	1.15 (0.98, 1.36)	1.39 (1.17, 1.65)	1.82 (1.51, 2.19)	0.39	< 0.001
<i>Pancreatic cancer mortality (874 cases)</i>	1.00	1.35 (1.05, 1.72)	1.17 (0.92, 1.49)	1.08 (0.83, 1.42)	1.25 (0.89, 1.74)	0.64	0.64
<i>Liver cancer (137 cases)</i>	1.00	1.72 (0.89, 3.33)	1.37 (0.71, 2.61)	1.76 (0.88, 3.52)	0.62 (0.19, 1.94)	0.53	0.90
<i>Breast cancer mortality (1222 cases)</i>	1.00	0.97 (0.81, 1.18)	0.97 (0.81, 1.16)	0.92 (0.75, 1.14)	1.00 (0.77, 1.31)	0.50	0.21
<i>Postmenopausal breast cancer mortality (1089 cases)</i>	1.00	0.96 (0.78, 1.19)	0.94 (0.78, 1.14)	0.93 (0.74, 1.16)	1.05 (0.80, 1.39)	0.80	0.23
<i>Ovary cancer mortality (578 cases)</i>	1.00	0.72 (0.54, 0.97)	0.88 (0.68, 1.14)	0.82 (0.61, 1.10)	0.90 (0.62, 1.31)	0.73	0.98
<i>Endometrial Cancer (242 cases)</i>	1.00	1.11 (0.70, 1.75)	1.19 (0.78, 1.82)	1.12 (0.69, 1.82)	1.18 (0.63, 2.21)	0.17	0.68
<i>Prostate cancer mortality (572 cases)</i>	1.00	0.76 (0.59, 0.98)	0.76 (0.59, 0.98)	0.73 (0.54, 1.00)	0.50 (0.30, 0.85)	0.08	0.12
<b>Respiratory disease mortality</b>	1.00	0.90	0.84	1.04	1.32	< 0.001	0.04

<b>(2,359 cases)</b>		(0.77, 1.05)	(0.73, 0.97)	(0.89, 1.22)	(1.10, 1.58)		
<b>Neurological disease mortality</b>	1.00	0.91	0.80	0.69	0.60	0.17	0.01
<b>(530 cases)</b>		(0.68, 1.21)	(0.61, 1.04)	(0.50, 0.95)	(0.38, 0.94)		
<b>Diabetes mortality</b>	1.00	0.90	0.59	0.65	0.57	0.03	0.003
<b>(338 cases)</b>		(0.65, 1.24)	(0.43, 0.82)	(0.43, 0.97)	(0.32, 1.01)		
<b>Injury mortality</b>	1.00	0.88	0.91	0.84	0.98	0.58	0.44
<b>(654 cases)</b>		(0.68, 1.13)	(0.71, 1.16)	(0.62, 1.12)	(0.67, 1.44)		
<b>Suicide</b>	1.00	1.10	0.76	0.58	0.54	0.70	< 0.001
<b>(374 cases)</b>		(0.81, 1.48)	(0.56, 1.03)	(0.39, 0.86)	(0.31, 0.92)		
<b>Other disease mortality</b>	1.00	0.88	0.80	0.81	0.79	< 0.001	< 0.001
<b>(8,272 cases)</b>		(0.82, 0.95)	(0.75, 0.86)	(0.74, 0.87)	(0.71, 0.88)		

\* A likelihood ratio test was performed.

The model adjusted for age, baseline disease status (hypertension, hypercholesterolemia, diabetes), BMI (< 20.9, 21-22.9, 23-24.9, 25-29.9, 30-34.9,  $\geq 35$  kg/m<sup>2</sup>), physical activity (< 3, 3-8.9, 9-17.9, 18-26.9,  $\geq 27$  MET-h/wk), overall dietary pattern (AHEI score, in quintiles), total energy intake (quintiles), smoking status (never, former (1 - 4 cigarettes/d), former (5 - 14 cigarettes/d), former (15 - 24 cigarettes/d), former (25 - 34 cigarettes/d), former (35 - 44 cigarettes/d), former ( $\geq 45$  cigarettes/d), former (unknown cigarettes/d), current (1 - 4 cigarettes/d), current (5 - 14 cigarettes/d), current (15 - 24 cigarettes/d), current (25 - 34 cigarettes/d), current (35 - 44 cigarettes/d), current ( $\geq 45$  cigarettes/d), current (unknown cigarettes/d)), sugar-sweetened beverages consumption (quintiles) and alcohol consumption (0, 0-5, 5-10, 10-15,  $\geq 15$  g/d). We additionally adjusted for menopausal status (yes vs. no), and postmenopausal hormone use (yes vs. no) for women. Caffeinated and decaffeinated coffee adjusted for each other.

**Supplemental Table 4. HRs (95% CI) for the association between consumption of total coffee and risk of cause-specific mortality among ever smokers**

	0 cup/d	<1 cup/d	1-3 cups/d	3-5 cups/d	>5 cups/d	P <sub>non-linearity*</sub>	P for linear trend
<b>CVD mortality</b>	1.00	1.03 (0.93, 1.14)	0.94 (0.85, 1.04)	0.93 (0.83, 1.04)	1.14 (1.00, 1.30)	0.005	0.11
<i>CHD mortality</i>	1.00	1.03 (0.91, 1.16)	0.95 (0.84, 1.07)	0.96 (0.84, 1.10)	1.17 (1.00, 1.36)	0.01	0.3
<i>Stroke mortality</i>	1.00	1.02 (0.82, 1.28)	0.91 (0.74, 1.12)	0.81 (0.64, 1.03)	1.05 (0.80, 1.38)	0.15	0.13
<b>Cancer mortality</b>	1.00	1.04 (0.95, 1.14)	1.07 (0.98, 1.16)	1.16 (1.06, 1.27)	1.41 (1.27, 1.57)	0.009	< 0.001
<i>Colorectal cancer mortality</i>	1.00	1.23 (0.89, 1.68)	1.04 (0.77, 1.41)	1.06 (0.76, 1.47)	1.02 (0.69, 1.51)	0.47	0.66
<i>Lung cancer mortality</i>	1.00	1.00 (0.82, 1.22)	1.19 (0.99, 1.42)	1.56 (1.29, 1.88)	2.37 (1.94, 2.89)	0.03	< 0.001
<i>Pancreatic cancer mortality</i>	1.00	1.23 (0.86, 1.76)	1.20 (0.85, 1.68)	1.11 (0.77, 1.61)	1.57 (1.04, 2.35)	0.92	0.63
<i>Liver cancer</i>	1.00	1.72 (0.89, 3.33)	1.37 (0.71, 2.61)	1.76 (0.88, 3.52)	0.62 (0.19, 1.94)	0.55	0.45
<i>Breast cancer mortality</i>	1.00	1.11 (0.80, 1.53)	1.09 (0.81, 1.47)	1.08 (0.78, 1.49)	1.32 (0.92, 1.90)	0.52	0.31
<i>Postmenopausal breast cancer mortality</i>	1.00	1.12 (0.79, 1.60)	1.14 (0.82, 1.58)	1.12 (0.79, 1.59)	1.42 (0.97, 2.09)	0.73	0.19
<i>Ovary cancer mortality</i>	1.00	0.81 (0.50, 1.33)	1.01 (0.66, 1.55)	1.12 (0.71, 1.75)	1.17 (0.70, 1.93)	0.71	0.12
<i>Endometrial Cancer</i>	1.00	1.28 (0.62, 2.67)	1.11 (0.56, 2.20)	1.15 (0.56, 2.39)	0.71 (0.28, 1.81)	0.22	0.27
<i>Prostate cancer mortality</i>	1.00	0.72 (0.51, 1.01)	0.67 (0.48, 0.94)	0.68 (0.46, 1.01)	0.42 (0.23, 0.78)	0.38	0.03

<b>Respiratory disease mortality</b>	1.00	0.88 (0.73, 1.05)	0.82 (0.69, 0.97)	1.07 (0.90, 1.29)	1.42 (1.16, 1.74)	< 0.001	< 0.001
<b>Neurological disease mortality</b>	1.00	0.90 (0.58, 1.39)	0.75 (0.50, 1.13)	0.68 (0.43, 1.06)	0.49 (0.27, 0.88)	0.77	0.02
<b>Diabetes mortality</b>	1.00	0.82 (0.53, 1.27)	0.52 (0.33, 0.80)	0.60 (0.36, 0.99)	0.54 (0.27, 1.06)	0.02	0.03
<b>Injury mortality</b>	1.00	0.79 (0.56, 1.13)	0.89 (0.64, 1.23)	0.87 (0.60, 1.25)	0.93 (0.59, 1.48)	0.41	0.75
<b>Suicide</b>	1.00	0.91 (0.60, 1.38)	0.75 (0.50, 1.11)	0.58 (0.36, 0.94)	0.58 (0.31, 1.06)	0.95	0.001
<b>Other disease mortality</b>	1.00	0.88 (0.79, 0.97)	0.79 (0.71, 0.87)	0.81 (0.73, 0.90)	0.80 (0.70, 0.91)	0.003	< 0.001

\* A likelihood ratio test was performed.

The model adjusted for age, baseline disease status (hypertension, hypercholesterolemia, diabetes), BMI (< 20.9, 21-22.9, 23-24.9, 25-29.9, 30-34.9,  $\geq 35$  kg/m<sup>2</sup>), physical activity (< 3, 3-8.9, 9-17.9, 18-26.9,  $\geq 27$  MET-h/wk), overall dietary pattern (AHEI score, in quintiles), total energy intake (quintiles), smoking status (never, former (1 - 4 cigarettes/d), former (5 - 14 cigarettes/d), former (15 - 24 cigarettes/d), former (25 - 34 cigarettes/d), former (35 - 44 cigarettes/d), former ( $\geq 45$  cigarettes/d), former (unknown cigarettes/d), current (1 - 4 cigarettes/d), current (5 - 14 cigarettes/d), current (15 - 24 cigarettes/d), current (25 - 34 cigarettes/d), current (35 - 44 cigarettes/d), current ( $\geq 45$  cigarettes/d), current (unknown cigarettes/d)), sugar-sweetened beverages consumption (quintiles) and alcohol consumption (0, 0-5, 5-10, 10-15,  $\geq 15$  g/d). We additionally adjusted for menopausal status (yes vs. no), and postmenopausal hormone use (yes vs. no) for women. Caffeinated and decaffeinated coffee adjusted for each other.

**Supplemental Table 5. Stratified analysis for the association between coffee consumption and risk of total mortality**

	Categories of total coffee consumption					P for interaction†
	0	<1 cup/d	1-3 cups/d	3-5 cups/d	>5 cups/d	
<b>Stratified by age</b>						0.003
<i>Age ≤ 70 years</i>						
Multivariate-adjusted hazard ratio (95% CI)	1.00	0.97 (0.92-1.03)	0.87 (0.83-0.92)	0.86 (0.81-0.92)	0.96 (0.88-1.04)	
<i>Age &gt; 70 years</i>						
Multivariate-adjusted hazard ratio (95% CI)	1.00	0.96 (0.91-1.01)	0.93 (0.89-0.98)	0.95 (0.89-1.00)	1.06 (0.99-1.14)	
<b>Stratified by BMI</b>						0.076
<i>BMI ≤ 25 kg/m<sup>2</sup></i>						
Multivariate-adjusted hazard ratio (95% CI)	1.00	0.93 (0.88-0.99)	0.92 (0.87-0.97)	0.93 (0.87-0.99)	1.03 (0.95-1.11)	
<i>BMI &gt; 25 kg/m<sup>2</sup></i>						
Multivariate-adjusted hazard ratio (95% CI)	1.00	0.97 (0.92-1.02)	0.90 (0.86-0.95)	0.91 (0.86-0.97)	0.98 (0.91-1.05)	
<b>Stratified by aHEI</b>						0.96
<i>aHEI ≤ median level</i>						
Multivariate-adjusted hazard ratio (95% CI)	1.00	0.95 (0.90-1.00)	0.90 (0.86-0.95)	0.92 (0.87-0.98)	1.00 (0.93-1.07)	
<i>aHEI &gt; median level</i>						
Multivariate-adjusted hazard ratio (95% CI)	1.00	0.95 (0.90-1.01)	0.92 (0.87-0.97)	0.94 (0.88-1.00)	1.04 (0.96-1.13)	
<b>Stratified by physical activity</b>						0.20
<i>Physical activity ≤ median level</i>						
Multivariate-adjusted hazard ratio (95% CI)	1.00	0.91 (0.87-0.96)	0.88 (0.84-0.92)	0.92 (0.87-0.97)	1.00 (0.93-1.07)	

<i>Physical activity &gt; median level</i>						
Multivariate-adjusted hazard ratio (95% CI)	1.00	1.01 (0.94-1.07)	0.96 (0.91-1.02)	0.95 (0.88-1.02)	1.06 (0.97-1.16)	< 0.001
<b>Stratified by smoking status</b>						
<i>Never</i>						
Multivariate-adjusted hazard ratio (95% CI)	1.00	0.94 (0.89-1.00)	0.92 (0.87-0.97)	0.85 (0.79-0.92)	0.88 (0.78-0.99)	0.39
<i>Ever</i>						
Multivariate-adjusted hazard ratio (95% CI)	1.00	0.96 (0.91-1.01)	0.92 (0.87-0.97)	0.96 (0.91-1.01)	1.04 (0.98-1.11)	
<b>Stratified by sex</b>						
<i>Female</i>						
Multivariate-adjusted hazard ratio (95% CI)	1.00	0.93 (0.88-0.97)	0.89 (0.85-0.93)	0.93 (0.88-0.98)	1.02 (0.95-1.08)	0.39
<i>Male</i>						
Multivariate-adjusted hazard ratio (95% CI)	1.00	1.00 (0.94, 1.06)	0.97 (0.91, 1.03)	0.96 (0.90, 1.03)	1.05 (0.96, 1.14)	
<b>Stratify by individual cohort (cohort effect)</b>						
NHS						
Multivariate-adjusted hazard ratio (95% CI)	1.00	0.94 (0.88, 1.00)	0.90 (0.85, 0.95)	0.92 (0.88, 0.98)	1.02 (0.96, 1.10)	0.39
NHS 2						
Multivariate-adjusted hazard ratio (95% CI)	1.00	0.92 (0.81, 1.04)	0.85 (0.75, 0.95)	0.86 (0.73, 1.01)	0.97 (0.78, 1.20)	
HPFS						



	1.00	1.00	0.97	0.96	1.05
Multivariate-adjusted hazard ratio (95% CI)		(0.94, 1.06)	(0.91, 1.03)	(0.90, 1.03)	(0.96, 1.14)

---

Abbreviations: CI, confidence interval; BMI, body mass index; aHEI, alternative health eating index

\* Models were adjusted for age, baseline disease status (hypertension, hypercholesterolemia, diabetes), BMI (< 20.9, 21-22.9, 23-24.9, 25-29.9, 30-34.9,  $\geq 35$  kg/m<sup>2</sup>), physical activity (< 3, 3-8.9, 9-17.9, 18-26.9,  $\geq 27$  MET-h/wk), overall dietary pattern (AHEI score, in quintiles), total energy intake (quintiles), smoking status (never, former (1 - 4 cigarettes/d), former (5 - 14 cigarettes/d), former (15 - 24 cigarettes/d), former (25 - 34 cigarettes/d), former (35 - 44 cigarettes/d), former ( $\geq 45$  cigarettes/d), former (unknown cigarettes/d), current (1 - 4 cigarettes/d), current (5 - 14 cigarettes/d), current (15 - 24 cigarettes/d), current (25 - 34 cigarettes/d), current (35 - 44 cigarettes/d), current ( $\geq 45$  cigarettes/d), current (unknown cigarettes/d)), sugar-sweetened beverages consumption (quintiles) and alcohol consumption (0, 0-5, 5-10, 10-15,  $\geq 15$  g/d). We additionally adjusted for menopausal status (yes vs. no), and postmenopausal hormone use (yes vs. no) for women.

\* † Likelihood ratio tests were performed.

**Supplemental Table 6. The tests of proportional hazard assumption in NHS, NHS 2, and HPFS.**

	Categories of total coffee consumption					P <sub>interaction</sub> <sup>†</sup>
	0	<1 cup/d	1-3 cups/d	3-5 cups/d	>5 cups/d	
<b>Stratified by follow-up time in NHS</b>						< 0.001
< 20 years						
Multivariate-adjusted hazard ratio (95% CI)	1.00	0.93 (0.85, 1.02)	0.87 (0.80, 0.94)	0.84 (0.76, 0.92)	0.89 (0.79, 0.99)	
≥ 20 years						
Multivariate-adjusted hazard ratio (95% CI)	1.00	0.94 (0.87, 1.01)	0.93 (0.87, 0.99)	1.00 (0.92, 1.07)	1.12 (1.02, 1.22)	
<b>Stratified by follow-up time in NHS2</b>						0.49
< 14 years						
Multivariate-adjusted hazard ratio (95% CI)	1.00	0.92 (0.74, 1.13)	0.99 (0.81, 1.21)	0.96 (0.74, 1.25)	1.04 (0.72, 1.48)	
≥ 14 years						
Multivariate-adjusted hazard ratio (95% CI)	1.00	0.91 (0.78, 1.06)	0.78 (0.67, 0.91)	0.83 (0.68, 1.00)	1.01 (0.78, 1.31)	
<b>Stratified by follow-up time in HPFS</b>						0.53
< 20 years						
Multivariate-adjusted hazard ratio (95% CI)	1.00	1.00 (0.92, 1.09)	0.96 (0.88, 1.04)	0.93 (0.85, 1.03)	1.00 (0.88, 1.14)	
≥ 20 years						
Multivariate-adjusted hazard ratio (95% CI)	1.00	0.98 (0.90, 1.07)	0.95 (0.87, 1.03)	0.96 (0.87, 1.06)	1.03 (0.90, 1.17)	

\* Models were adjusted for age, baseline disease status (hypertension, hypercholesterolemia, diabetes), BMI (< 20.9, 21-22.9, 23-24.9, 25-29.9, 30-34.9,  $\geq 35$  kg/m<sup>2</sup>), physical activity (< 3, 3-8.9, 9-17.9, 18-26.9,  $\geq 27$  MET-h/wk), overall dietary pattern (AHEI score, in quintiles), total energy intake (quintiles), smoking status (never, former (1 - 4 cigarettes/d), former (5 - 14 cigarettes/d), former (15 - 24 cigarettes/d), former (25 - 34 cigarettes/d), former (35 - 44 cigarettes/d), former ( $\geq 45$  cigarettes/d), former (unknown cigarettes/d), current (1 - 4 cigarettes/d), current (5 - 14 cigarettes/d), current (15 - 24 cigarettes/d), current (25 - 34 cigarettes/d), current (35 - 44 cigarettes/d), current ( $\geq 45$  cigarettes/d), current (unknown cigarettes/d)), sugar-sweetened beverages consumption (quintiles) and alcohol consumption (0, 0-5, 5-10, 10-15,  $\geq 15$  g/d). We additionally adjusted for menopausal status (yes vs. no), and postmenopausal hormone use (yes vs. no) for women.

\* † Likelihood ratio tests were performed.

**Supplemental Table 7. Sensitivity analyses for the association between coffee consumption and total mortality in the overall population, and pooled multivariate-adjusted hazard ratio was shown**

	Categories of total coffee					P <sub>non-linearity*</sub>
	0	<1 cup/d	1-3 cups/d	3-5 cups/d	>5 cups/d	
Using cumulative coffee consumption, and stopping updating after cancer and diabetes	1.00	0.96 (0.92, 1.00)	0.89 (0.86, 0.92)	0.91 (0.87, 0.95)	0.94 (0.88, 0.99)	< 0.001
Using cumulative coffee consumption, and further stopping updating after hypertension, hypercholesterolemia, and CVD	1.00	0.96 (0.92, 1.00)	0.89 (0.86, 0.92)	0.91 (0.87, 0.95)	0.97 (0.92, 1.03)	< 0.001
Continue updating after diagnosis of chronic disease with 4-year lag	1.00	0.99 (0.96, 1.04)	0.93 (0.88, 0.97)	0.94 (0.89, 0.98)	0.94 (0.87, 1.00)	< 0.001
Continue updating after diagnosis of chronic disease adjusting for time-varying hypercholesterolemia	1.00	1.02 (0.97, 1.06)	0.92 (0.88, 0.96)	0.92 (0.87, 0.96)	0.86 (0.80, 0.92)	0.04
Using baseline exposure with exclusion of hypertension and hypercholesterolemia at baseline	1.00	0.98 (0.93, 1.02)	0.93 (0.88, 0.97)	0.95 (0.91, 1.00)	1.04 (0.98, 1.10)	< 0.001

	Categories of caffeinated coffee					
	0	<1 cup/d	1-3 cups/d	3-5 cups/d	>5 cups/d	
Using cumulative coffee consumption, and stopping updating after cancer and diabetes	1.00	0.98 (0.95, 1.01)	0.93 (0.90, 0.96)	0.98 (0.94, 1.03)	0.98 (0.91, 1.04)	< 0.001
Using cumulative coffee consumption, and further stopping updating after hypertension, hypercholesterolemia, and CVD	1.00	0.98 (0.95, 1.01)	0.91 (0.88, 0.94)	0.98 (0.94, 1.03)	1.02 (0.96, 1.09)	< 0.001
Continue updating after diagnosis of chronic disease with 4-year lag		1.00	0.95	0.99	0.98	0.014
	1.00	(0.97, 1.04)	(0.92, 0.99)	(0.95, 1.03)	(0.91, 1.07)	
Continue updating after diagnosis of chronic disease adjusting for time-varying hypercholesterolemia		1.00	0.94	0.95	0.87	< 0.001
	1.00	(0.97, 1.04)	(0.91, 0.97)	(0.91, 1.00)	(0.80, 0.95)	
Using baseline exposure with exclusion of hypertension and hypercholesterolemia at baseline	1.00	0.98 (0.95, 1.01)	0.93 (0.90, 0.96)	1.00 (0.96, 1.06)	1.10 (1.03, 1.14)	< 0.001
	Categories of decaffeinated coffee					
	0	<1 cup/d	1-3 cups/d	>3 cups/d		

Using cumulative coffee consumption, and stopping updating after cancer and diabetes	1.00	0.89 (0.86, 0.91)	0.87 (0.84, 0.90)	0.94 (0.88, 0.99)	< 0.001
Using cumulative coffee consumption, and further stopping updating after hypertension, hypercholesterolemia, and CVD	1.00	0.86 (0.86, 0.88)	0.84 (0.82, 0.88)	0.95 (0.88, 1.01)	< 0.001
Continue updating after diagnosis of chronic disease with 4-year lag	1.00	0.90 (0.88, 0.92)	0.89 (0.87, 0.93)	0.91 (0.85, 0.98)	< 0.001
Continue updating after diagnosis of chronic disease adjusting for time-varying hypercholesterolemia	1.00	0.90 (0.87, 0.92)	0.88 (0.86, 0.92)	0.90 (0.84, 0.97)	< 0.001
Using baseline exposure with exclusion of hypertension and hypercholesterolemia at baseline	1.00	0.93 (0.89, 0.96)	0.93 (0.89, 0.97)	0.93 (0.86, 1.00)	< 0.001

Models were adjusted for age, baseline disease status (hypertension, hypercholesterolemia, diabetes), BMI (< 20.9, 21-22.9, 23-24.9, 25-29.9, 30-34.9,  $\geq 35$  kg/m<sup>2</sup>), physical activity (< 3, 3-8.9, 9-17.9, 18-26.9,  $\geq 27$  MET-h/wk), overall dietary pattern (AHEI score, in quintiles), total energy intake (quintiles), smoking status (never, former (1 - 4 cigarettes/d), former (5 - 14 cigarettes/d), former (15 - 24 cigarettes/d), former (25 - 34 cigarettes/d), former (35 - 44 cigarettes/d), former ( $\geq 45$  cigarettes/d), former (unknown cigarettes/d), current (1 - 4 cigarettes/d), current (5 - 14 cigarettes/d), current (15 - 24 cigarettes/d), current (25 - 34 cigarettes/d), current (35 - 44

cigarettes/d), current ( $\geq 45$  cigarettes/d), current (unknown cigarettes/d)), sugar-sweetened beverages consumption (quintiles) and alcohol consumption (0, 0-5, 5-10, 10-15,  $\geq 15$  g/d). We additionally adjusted for menopausal status (yes vs. no), and postmenopausal hormone use (yes vs. no) for women.

\* A likelihood ratio test was performed.

**Supplemental Table 8. Sensitivity analyses for the association between coffee consumption and total mortality among never smokers, and pooled multivariate-adjusted hazard ratio was shown.**

	Categories of total coffee					P <sub>non-linearity*</sub>
	0	<1 cup/d	1-3 cups/d	3-5 cups/d	>5 cups/d	
Using cumulative coffee consumption, and stopping updating after cancer and diabetes	1.00	0.93 (0.87, 0.99)	0.89 (0.83, 0.95)	0.83 (0.76, 0.90)	0.83 (0.72, 0.96)	0.26
Using cumulative coffee consumption, and further stopping updating after hypertension, hypercholesterolemia, and CVD	1.00	0.93 (0.87, 0.99)	0.89 (0.83, 0.95)	0.84 (0.78, 0.91)	0.85 (0.74, 0.97)	0.87
Continue updating after diagnosis of chronic disease with 4-year lag	1.00	0.96 (0.90, 1.03)	0.93 (0.88, 1.00)	0.80 (0.73, 0.86)	0.85 (0.72, 1.00)	0.25
Continue updating after diagnosis of chronic disease adjusting for time-varying hypercholesterolemia	1.00	0.99 (0.93, 1.06)	0.91 (0.85, 0.96)	0.79 (0.73, 0.85)	0.73 (0.61, 0.87)	0.30
Using baseline exposure with exclusion of hypertension and hypercholesterolemia at baseline	1.00	0.95 (0.87, 1.01)	0.93 (0.88, 1.00)	0.82 (0.75, 0.89)	0.91 (0.79, 1.04)	0.88



	Categories of caffeinated coffee					
	0	<1 cup/d	1-3 cups/d	3-5 cups/d	>5 cups/d	
Using cumulative coffee consumption, and stopping updating after cancer and diabetes	1.00	1.00 (0.95, 1.05)	0.91 (0.86, 0.96)	0.87 (0.79, 0.95)	0.80 (0.65, 0.99)	0.34
Using cumulative coffee consumption, and further stopping updating after hypertension, hypercholesterolemia, and CVD	1.00	1.00 (0.93, 1.05)	0.88 (0.83, 0.93)	0.89 (0.82, 0.98)	0.79 (0.66, 0.97)	0.91
Continue updating after diagnosis of chronic disease with 4-year lag	1.00	0.99 (0.94, 1.05)	0.90 (0.85, 0.96)	0.85 (0.76, 0.94)	0.81 (0.63, 1.06)	0.084
Continue updating after diagnosis of chronic disease adjusting for time-varying hypercholesterolemia	1.00	0.99 (0.94, 1.05)	0.89 (0.84, 0.94)	0.80 (0.72, 0.88)	0.77 (0.59, 1.01)	0.07
Using baseline exposure with exclusion of hypertension and hypercholesterolemia at baseline	1.00	0.97 (0.92, 1.02)	0.92 (0.87, 0.99)	0.85 (0.76, 0.94)	0.94 (0.78, 1.12)	0.91
	Categories of decaffeinated coffee					
	0	<1 cup/d	1-3 cups/d	>3 cups/d		

Using cumulative coffee consumption, and stopping updating after cancer and diabetes	1.00	0.92 (0.87, 0.96)	0.90 (0.84, 0.95)	0.93 (0.81, 1.08)	0.016
Using cumulative coffee consumption, and further stopping updating after hypertension, hypercholesterolemia, and CVD	1.00	0.90 (0.86, 0.95)	0.89 (0.84, 0.95)	0.97 (0.85, 1.12)	0.008
Continue updating after diagnosis of chronic disease with 4-year lag	1.00	0.95 (0.90, 0.99)	0.97 (0.91, 1.03)	0.91 (0.76, 1.09)	0.22
Continue updating after diagnosis of chronic disease adjusting for time-varying hypercholesterolemia	1.00	0.94 (0.90, 0.98)	0.92 (0.87, 0.98)	0.89 (0.75, 1.06)	0.20
Using baseline exposure with exclusion of hypertension and hypercholesterolemia at baseline	1.00	0.95 (0.90, 0.99)	0.93 (0.87, 0.99)	0.87 (0.74, 1.04)	0.17

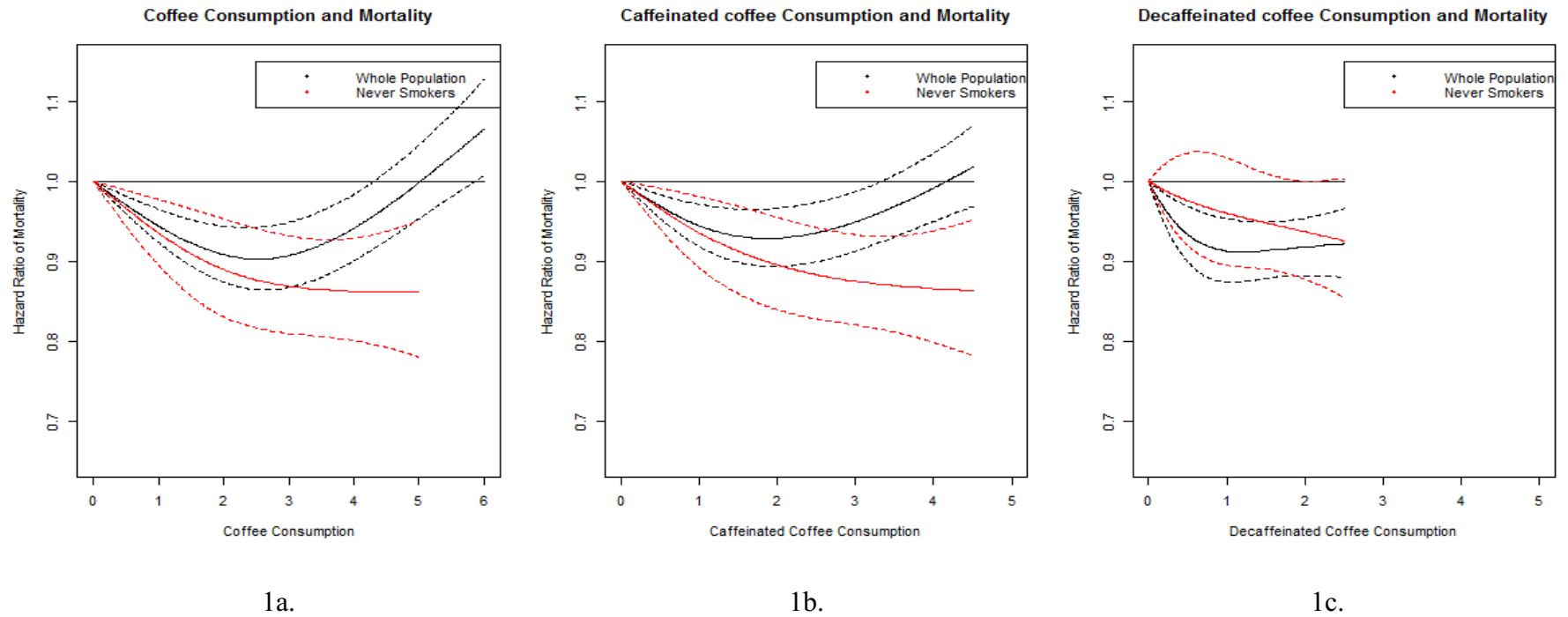
\* A likelihood ratio test was performed.

**Supplemental Table 9. The association between coffee consumption and risk of total mortality among never smokers by Cox model with inverse probability weighting**

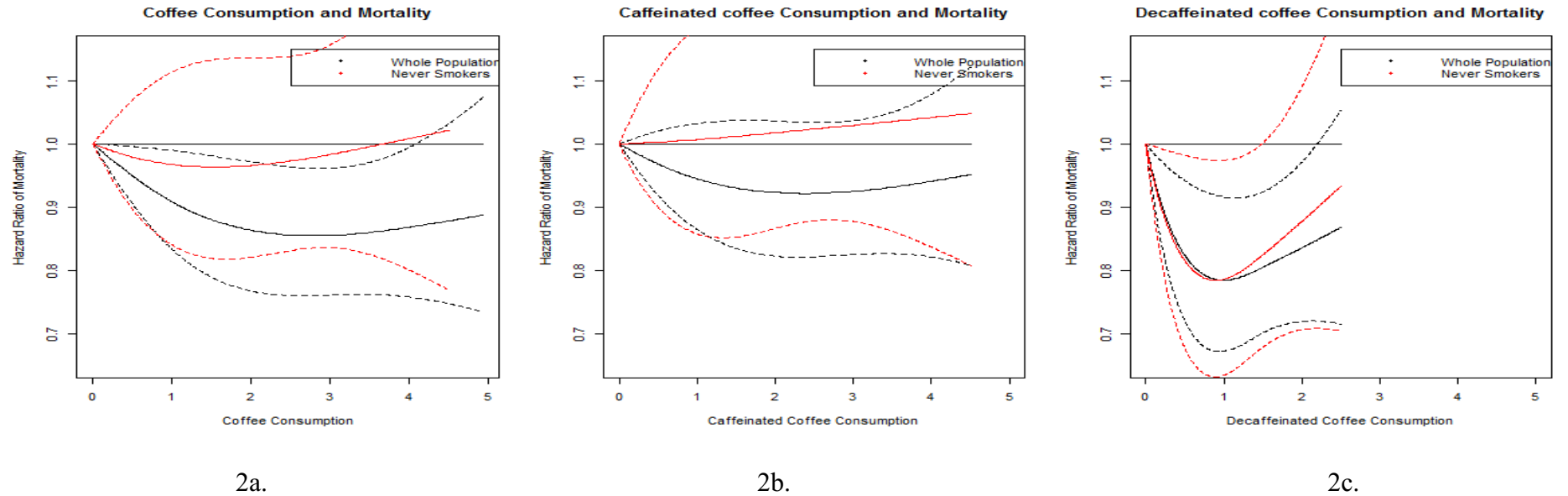
<b>Total coffee</b>					
Intake (cups/d)	0	<1 cup/d	1-3 cups/d	3-5 cups/d	>5 cups/d
Multivariate-adjusted hazard ratio (95% CI)	1.00	0.95 (0.90, 1.00)	0.93 (0.87, 0.98)	0.86 (0.80, 0.93)	0.87 (0.77, 0.98)
<b>Caffeinated coffee</b>					
Intake (cups/d)	0	<1 cup/d	1-3 cups/d	3-5 cups/d	>5 cups/d
Multivariate-adjusted hazard ratio (95% CI)	1.00	0.97 (0.94, 1.02)	0.92 (0.87, 0.96)	0.90 (0.82, 0.98)	0.89 (0.76, 1.05)
<b>Decaffeinated coffee</b>					
Intake (mg/d)	0	<1 cup/d	1-3 cups/d	>3 cups/d	
Multivariate-adjusted hazard ratio (95% CI)	1.00	0.95 (0.91, 0.99)	0.93 (0.87, 0.99)	0.89 (0.78, 1.03)	

The model adjusted for age, baseline disease status (hypertension, hypercholesterolemia, diabetes), BMI (< 20.9, 21-22.9, 23-24.9, 25-29.9, 30-34.9,  $\geq 35$  kg/m<sup>2</sup>), physical activity (< 3, 3-8.9, 9-17.9, 18-26.9,  $\geq 27$  MET-h/wk), overall dietary pattern (AHEI score, in quintiles), total energy intake (quintiles), sugar-sweetened beverages consumption (quintiles) and alcohol consumption (0, 0-5, 5-10, 10-15,  $\geq 15$  g/d). We additionally adjusted for menopausal status (yes vs. no), and postmenopausal hormone use (yes vs. no) for women. Caffeinated and decaffeinated coffee adjusted for each other.

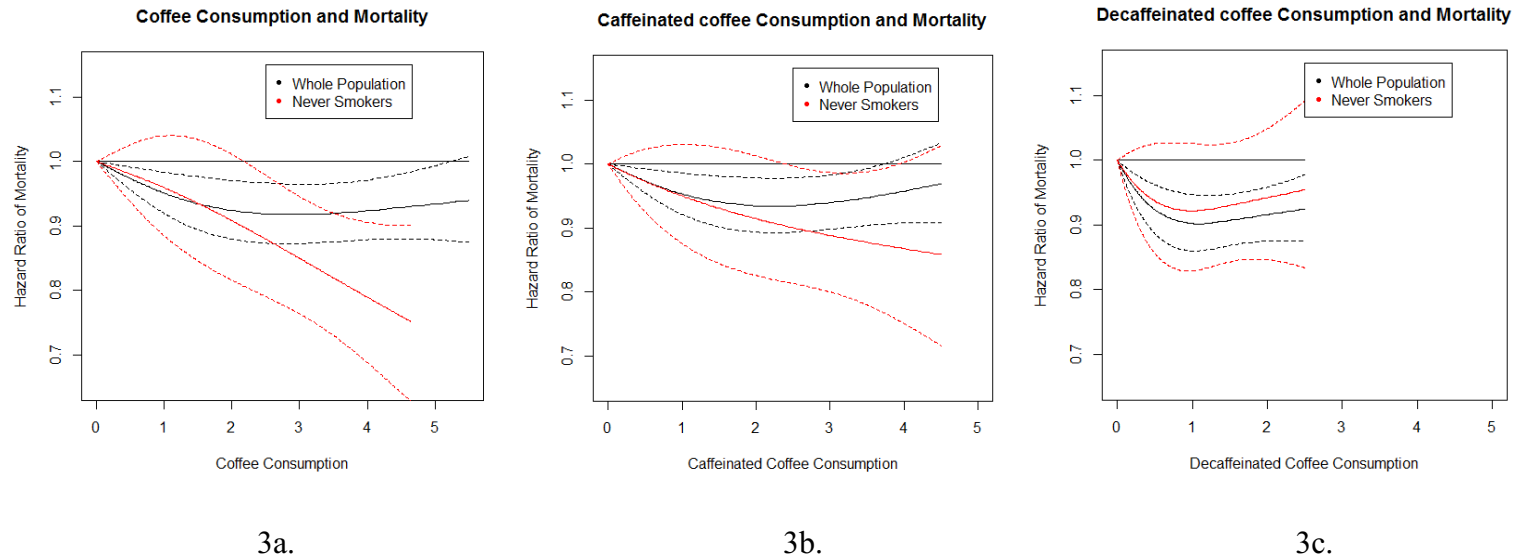
For the weight calculation, the weight for the non-cases was 1, while the weights for the death cases were calculated in a way that the composition of the total mortality among the never smokers was the same as the overall population.



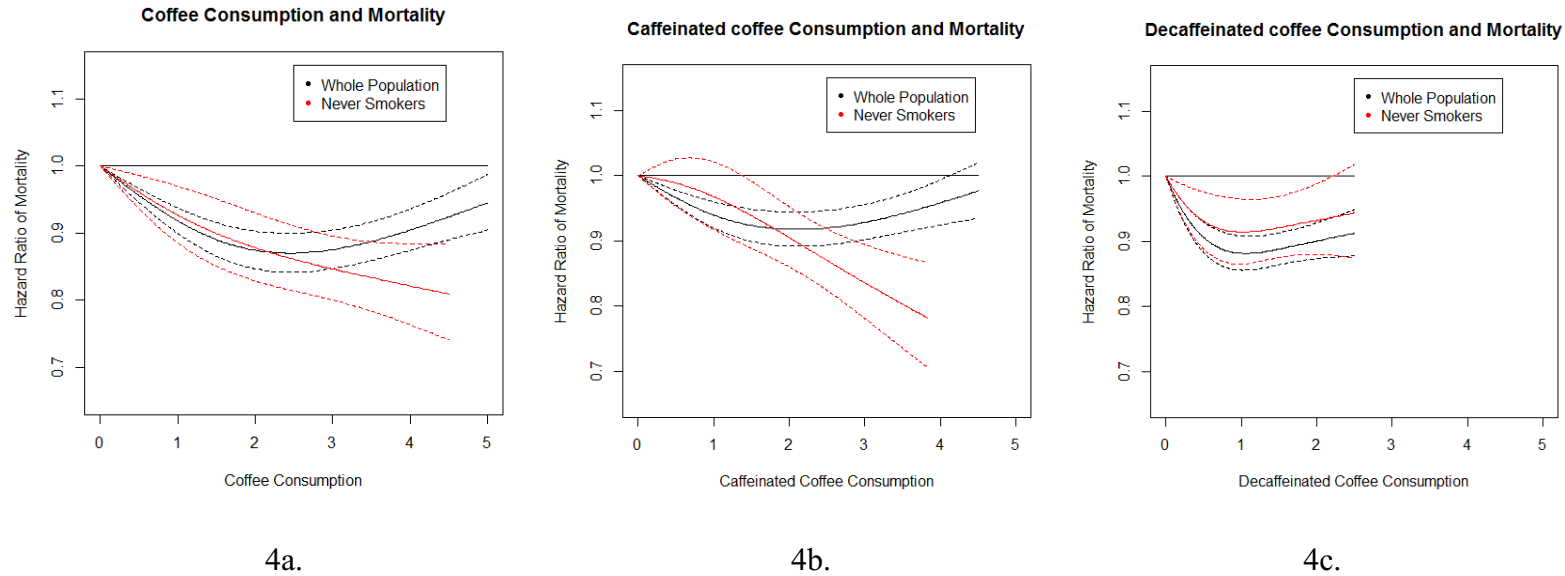
**Supplemental Figure 1. Baseline coffee consumption and risk of mortality in the overall population and among never smokers in NHS. 2a. Total coffee consumption and risk of mortality. 2b. Caffeinated coffee consumption and risk of mortality. 2c. Decaffeinated coffee consumption and risk of mortality.**



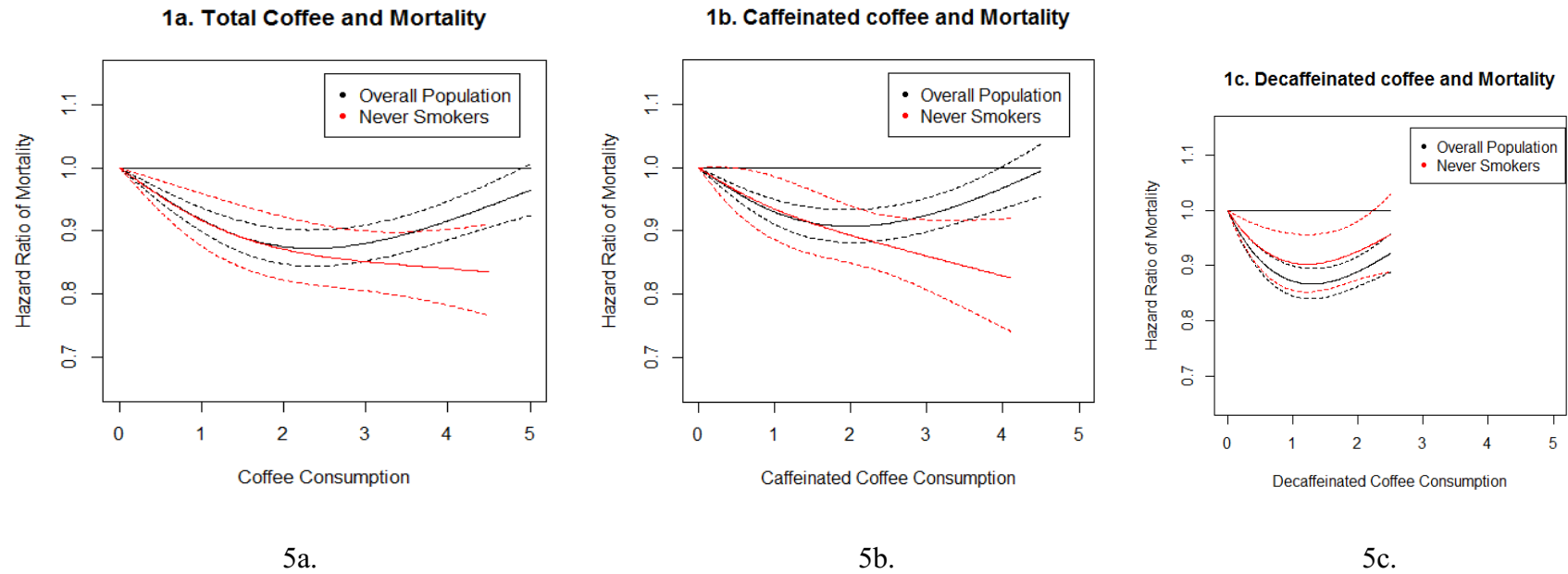
**Supplemental Figure 2. Baseline coffee consumption and risk of mortality in the overall population and among never smokers in NHS 2. 2a. Total coffee consumption and risk of mortality. 2b. Caffeinated coffee consumption and risk of mortality. 2c. Decaffeinated coffee consumption and risk of mortality.**



**Supplemental Figure 3. Baseline coffee consumption and risk of mortality in the overall population and among never smokers in HPFS. 3a. Total coffee consumption and risk of mortality. 3b. Caffeinated coffee consumption and risk of mortality. 3c. Decaffeinated coffee consumption and risk of mortality.**

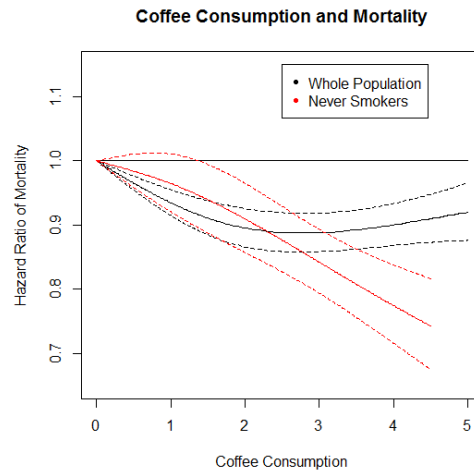


**Supplemental Figure 4. Cumulative coffee consumption and stopping updating when cancer and diabetes develop, and risk of mortality in the overall population and among never smokers by pooled across the three cohorts. 4a. Total coffee consumption and risk of mortality. 4b. Caffeinated coffee consumption and risk of mortality. 4c. Decaffeinated coffee consumption and risk of mortality.**

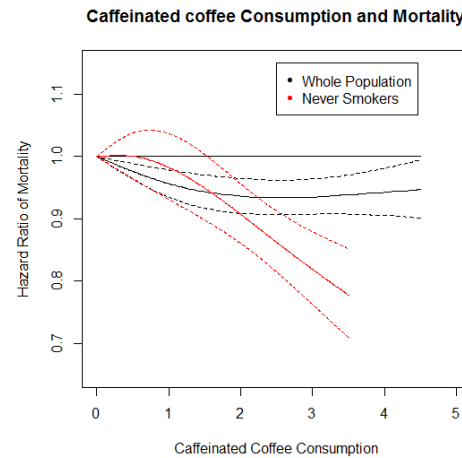


**Supplemental Figure 5. Cumulative coffee consumption and further stopping updating when hypertension, hypercholesterolemia develop, and risk of mortality in the overall population and among never smokers by pooled across the three cohorts. 5a. Total coffee consumption and risk of mortality. 5b. Caffeinated coffee consumption and risk of mortality. 5c. Decaffeinated coffee consumption and risk of mortality.**

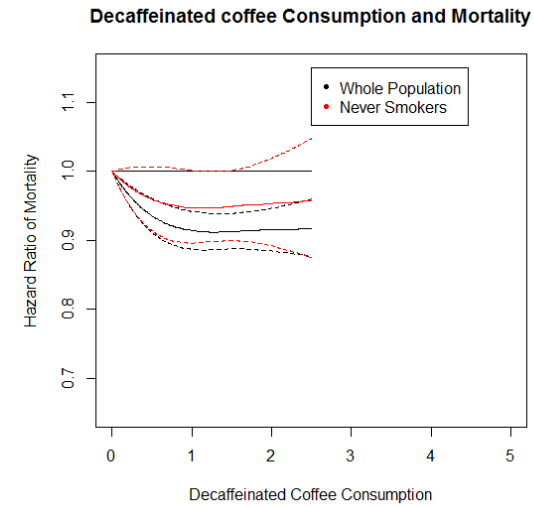




6a.

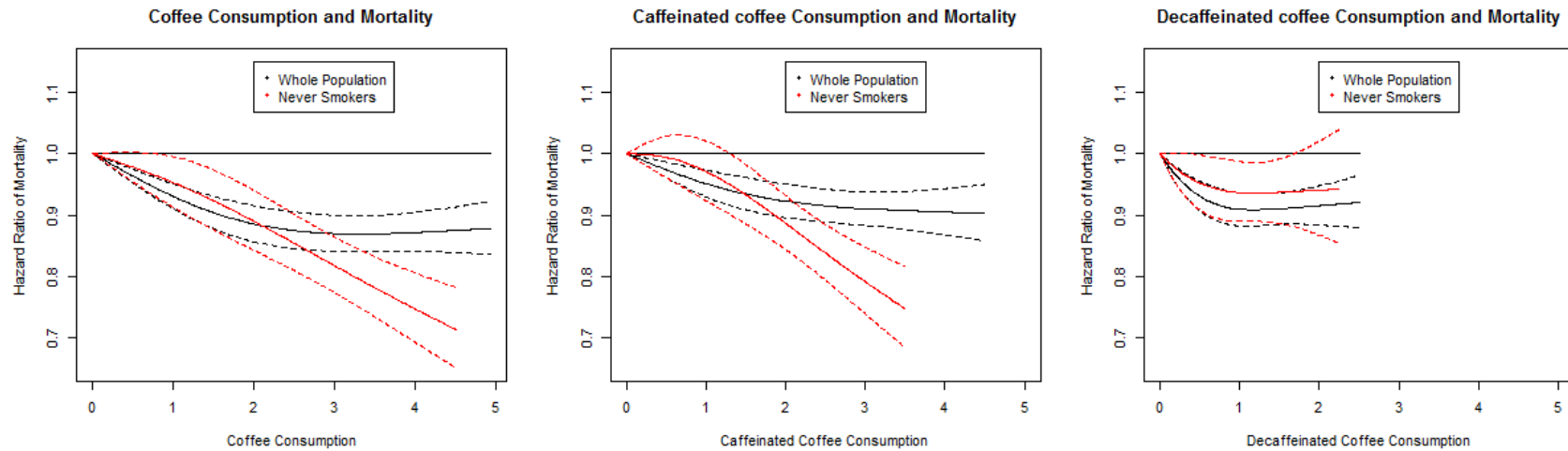


6b.



6c.

**Supplemental Figure 6. Continue updating coffee consumption after diagnosis of chronic disease with 4-year lag and risk of mortality in the overall population and among never smokers pooled across the three cohorts. 6a. Total coffee consumption and risk of mortality. 6b. Caffeinated coffee consumption and risk of mortality. 6c. Decaffeinated coffee consumption and risk of mortality.**

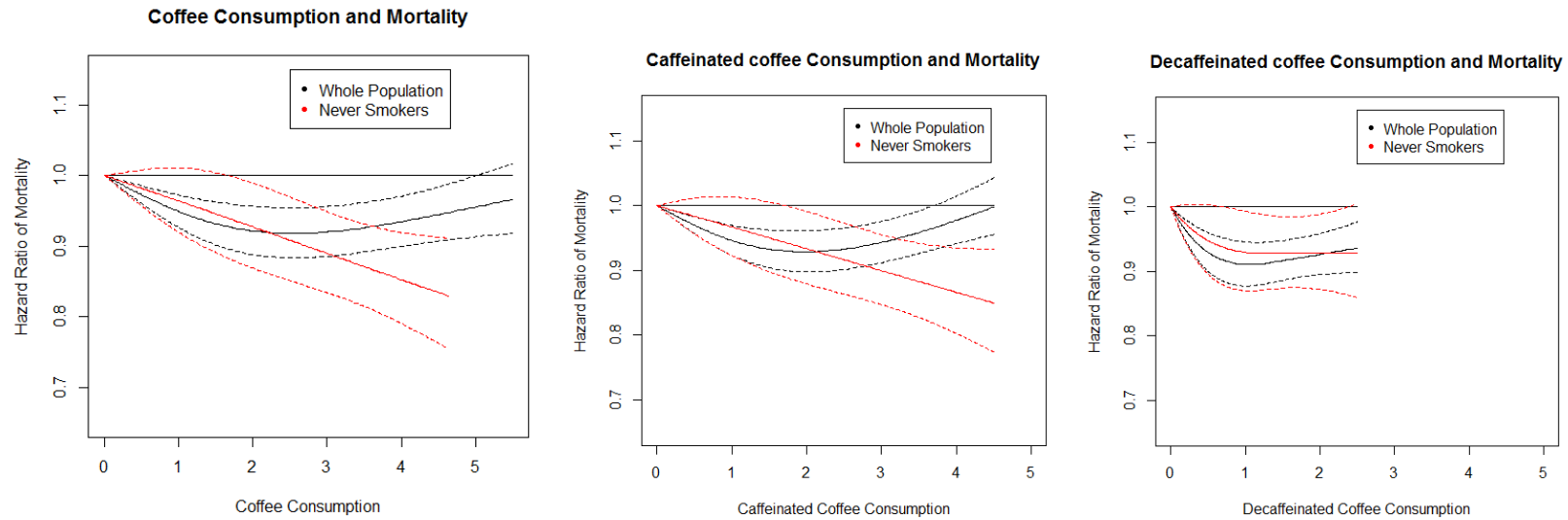


7a.

7b.

7c.

**Supplemental Figure 7. Continue updating coffee consumption after diagnosis of chronic disease adjusting for hypercholesterolemia as a time varying covariates and risk of mortality in the overall population and among never smokers pooled across the three cohorts. 7a. Total coffee consumption and risk of mortality. 7b. Caffeinated coffee consumption and risk of mortality. 7c. Decaffeinated coffee consumption and risk of mortality.**



8a.

8b.

8c.

**Supplemental Figure 8. Baseline coffee consumption and risk of mortality in the overall population and among never smokers further excluding hypertension, hypercholesterolemia, and diabetes cases at baseline, pooled across the three cohorts. 8a. Total coffee consumption and risk of mortality. 8b. Caffeinated coffee consumption and risk of mortality. 8c. Decaffeinated coffee consumption and risk of mortality.**

## CHAPTER 3

---

### CONSUMPTION OF SOY FOODS AND ISOFLAVONES AND RISK

### OF TYPE 2 DIABETES: A POOLED ANALYSIS OF THREE US

### COHORTS

---

Ming Ding <sup>1</sup>, An Pan <sup>2</sup>, JoAnn E. Manson <sup>3,4</sup>, Walter C. Willett <sup>1,4,5</sup>, Vasanti Malik <sup>1</sup>,  
Bernard Rosner <sup>5,6</sup>, Edward Giovannucci <sup>1,4,5</sup>, Frank B. Hu <sup>1,4,5</sup>, Qi Sun <sup>1,5</sup>

<sup>1</sup>Department of Nutrition, Harvard School of Public Health, Boston, MA

<sup>2</sup>Department of Epidemiology and Biostatistics, MOE Key Lab of Environment and Health, School of Public Health, Tongji Medical College, Huazhong University of Science and Technology, Wuhan 430030, Hubei, China

<sup>3</sup> Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA

<sup>4</sup>Department of Epidemiology, Harvard School of Public Health, Boston, MA

<sup>5</sup> Channing Division of Network Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA

<sup>6</sup>Department of Statistics, Harvard School of Public Health, Boston, MA

**ABSTRACT**

*Background* Evidence regarding the consumption of soy foods and isoflavones in relation to risk of type 2 diabetes (T2D) is scarce.

*Objective* To evaluate the association between soy food and isoflavone consumption and risk of T2D in US men and women.

*Methods* We followed 63,115 women in the Nurses' Health Study (1998-2012), 79,061 women in the Nurses' Health Study II (1999-2013), and 21,281 men in the Health Professionals Follow-Up Study (2002-2010). Diet was assessed by a validated food-frequency questionnaire, and was updated every 4 y. Self-reports of incident T2D was confirmed by a validated supplementary questionnaire.

*Results* During 1,966,321 person-years of follow-up, 9,185 incident T2D cases were documented. After multivariate adjustment for covariates, consumption of soy foods (tofu and soy milk) was not associated with a lower T2D risk. Compared to non-consumers of soy foods, the hazard ratio (HR) was 1.00 (95% CI: 0.93, 1.07) for <1 serving/week, and 0.93 (95% CI: 0.83, 1.03) for  $\geq 1$  serving/week of soy foods (P for trend = 0.14). In contrast, intake of total isoflavones was inversely associated with T2D risk. Comparing extreme quintiles of isoflavones, the HR was 0.89 (95% CI: 0.83, 0.96; P for trend = 0.009). Inverse associations were also found for consumption of major individual isoflavones, including daidzein and genistein, with risk of T2D.

*Conclusions* Intake of isoflavones was associated with a modestly lower T2D risk in US men and women who typically consumed low to moderate amounts of soy foods. These findings warrant replications in other populations with similar soy intake levels.

## INTRODUCTION

Type 2 diabetes (T2D) is a chronic disease with increasing prevalence worldwide. The total number of people with diabetes, globally, is estimated to reach 592 million by the year 2035.<sup>1</sup> Identification of modifiable lifestyle and dietary risk factors for T2D prevention is of high priority. Specific components of plant-based foods have been shown to exert significant health benefits.<sup>2</sup> For example, consumption of coffee and blueberries has been associated with a lower risk of T2D in Western populations, and certain flavonoid subclasses, such as phenolic acids and anthocyanins, may contribute to the health benefits of these foods.<sup>3-5</sup> In contrast, evidence regarding other plant-based foods, such as soy foods, that are regarded as healthful but not intrinsic to the traditional Western diet is sparse.

Soy foods are uniquely rich in isoflavones compared to other foods.<sup>6</sup> Isoflavones have a structure analogous to 17- $\beta$ -estradiol and have weak estrogen-like effects by binding to estrogen receptors.<sup>7</sup> Several clinical trials have been conducted to examine the effects of soy foods and isoflavones on glucose homeostasis, and results have suggested that soy foods and soy-rich diets may lower blood glucose.<sup>8-11</sup> However, these clinical trials are limited by small sample sizes and short durations of follow-up. Few prospective studies have been conducted to evaluate the associations between intakes of soy foods and isoflavones and T2D risk in Western populations who consume low to moderate amounts of soy foods<sup>12</sup>.

We conducted a prospective analysis of data collected in 3 large US cohorts, the Nurses' Health Study (NHS), the NHSII, and the Health Professionals Follow-Up Study

(HPFS) to examine the associations between consumption of soy food and isoflavones and risk of T2D.

## **METHODS**

### **Study population**

The NHS began in 1976, when 121,700 female registered nurses aged 30-55 y residing in 11 states were enrolled and completed a baseline questionnaire about their lifestyle and medical history. The NHSII was established in 1989 and consisted of 116,671 younger female registered nurses aged 25-42 y at baseline. These women responded to a baseline questionnaire similar to the one used in NHS. The HPFS was initiated in 1986, and was composed of 51,529 male dentists, pharmacists, veterinarians, optometrists, osteopathic physicians, and podiatrists aged 40-75 y at baseline. The male participants returned a baseline questionnaire about their medical history, lifestyle, and usual diet. In all three cohorts, questionnaires were administered at baseline and biennially thereafter to update information on lifestyle factors and the occurrence of chronic diseases.

For the current analysis, we excluded participants who reported diagnosed diabetes (including type 1 and type 2 diabetes, and gestational diabetes), cardiovascular disease (CVD), or cancer at baseline (1998 for the NHS, 1999 for the NHSII, and 2002 for the HPFS). We further excluded participants with missing soy or isoflavone consumption at baseline (when soy milk was first included) and those who left more than 70 food items blank or had daily energy intakes <600 or > 3500 kcal for women and <800 or >4200 kcal for men. Overall, 21,665 NHS participants, 8,537 NHSII participants,

and 16,120 HPFS participants were excluded from the analysis. After these exclusions, data from 63,115 NHS participants, 79,061 NHSII participants, and 21,281 HPFS participants were available for the analysis. The study protocol was approved by the institutional review boards of Brigham and Women's Hospital and Harvard School of Public Health. The completion of the self-administered questionnaire was considered to imply informed consent.

### **Assessment of isoflavone and soy food consumption**

In 1984, a 116-item food frequency questionnaire (FFQ) was administered to the NHS participants to obtain information on usual intake of food and beverages. Since 1986, an expanded FFQ has been administered every 4 years to update diet. Using a similar FFQ, dietary data were collected every four years from the NHSII participants since 1991 and from the HPFS participants since 1986. In all FFQs, participants were asked how often (from “never or less than once per month” to “6 or more times per day”) on average they consumed each food item of a standard portion size during the previous year. Major soy foods, i.e., tofu and soy milk, have been simultaneously included on the FFQs since 1998 in the NHS, 1999 in the NHSII, and 2002 in the HPFS. We therefore used these years as study baselines. Intake of isoflavones and other nutrients was calculated by multiplying the consumption frequency of each food item by the nutrient content of the specified portion and summing the contributions from all food items. We calculated consumption of genistein, daidzein, and glycitein from foods. Isoflavones from supplements were not included in these calculations. The food composition of isoflavones was created primarily from the USDA Database for the Isoflavone Content of Selected Foods, Release 2.0<sup>6</sup>. Consumption of total soy food was calculated as the sum of the consumption of tofu and



soy milk in servings/day. The validity and reproducibility of the FFQ has been described in detail elsewhere.<sup>13-16</sup> The correlation coefficient for tofu consumption assessed by FFQs and diet records was 0.56.<sup>13</sup>

### **Assessment of covariates**

In the biennial follow-up questionnaires, we collected and updated information on age, body weight and height, smoking status, physical activity, medication use, family history of diabetes, and disease status, including hypertension, hypercholesterolemia, CVD, and cancer. We also ascertained data on menopausal status and postmenopausal hormone use in both NHS and NHSII, as well as oral contraceptive use in NHSII. An overall measurement of diet quality was derived using the alternate Healthy Eating Index (aHEI) score excluding tofu and soy milk.

### **Assessment of type 2 diabetes (T2D)**

Participants with self-reported incident T2D were mailed a validated supplementary questionnaire regarding symptoms, diagnostic tests, and hypoglycemic therapy to confirm the diagnosis of diabetes. Cases were ascertained using the American Diabetes Association criteria:<sup>17</sup> 1) one or more classic symptoms (excessive thirst, polyuria, weight loss, hunger) and fasting plasma glucose concentrations  $\geq 7.0$  mmol/L or random plasma glucose concentrations  $\geq 11.1$  mmol/L; 2)  $\geq 2$  elevated plasma glucose concentrations on different occasions (fasting concentrations  $\geq 7.0$  mmol/L, random plasma glucose concentrations  $\geq 11.1$  mmol/L, and/or concentrations of  $\geq 11.1$  mmol/L after  $\geq 2$  h shown by oral-glucose-tolerance testing) in the absence of symptoms; or 3) treatment with hypoglycemic medication (insulin or oral hypoglycemic agent). In addition, hemoglobin

A1c $\geq$ 6.5% was added to the diagnosis criteria since 2010. Only cases confirmed by the supplemental questionnaires were included in our analysis.

The validity of the supplementary questionnaire for the diagnosis of diabetes has been documented previously. In a validation study, of the 62 cases in the NHS and 59 cases in HPFS who were confirmed by the supplemental questionnaire, 61 (98%) and 57 (97%) were reconfirmed by reviewing medical records.<sup>18,19</sup>

### **Statistical analysis**

We calculated person-time for each individual from the date of the return of the baseline questionnaire to the date of diagnosis of T2D, death, or the end of follow-up (30 June 2012 for the NHS, 30 June 2013 for the NHSII, and 31 January 2010 for the HPFS), whichever came first. We used cumulative averages of soy food or isoflavone consumption to reflect long-term dietary habits. We stopped updating diet after incident cancer or CVD as these diseases may result in changes of diet that might confound the association between soy foods and risk of T2D. In addition, to minimize missing values during follow-up, we replaced missing soy food/isoflavone intakes during follow-up with valid values in the previous cycle. We used Cox proportional hazards regression models to examine the associations between soy foods and isoflavone consumption (quintiles) and risk of T2D. The regression models included calendar time in 2-y intervals as the time scale, and were stratified by age in years. The pooled hazard ratios (HRs) were estimated by a stratified Cox model, which allowed baseline hazard to be different across the three cohorts while gave common effect estimates of the covariates. In multivariable analysis, we further adjusted for race (Caucasian, African American, Asian, and others), family history of T2D (yes, no), baseline disease status (hypertension and

hypercholesterolemia), BMI (<20.9, 21-22.9, 23-24.9, 25-29.9, 30-34.9,  $\geq 35$  kg/m<sup>2</sup>), physical activity (quintiles, met-hr/week), aHEI score (in quintiles), total energy intake (quintiles), and smoking status (never smoked, past smoker, currently smoked 1-14 cigarettes/d, and currently smoked >14 cigarettes/d). We additionally adjusted for menopausal status (yes, no), and postmenopausal hormone use (yes, no) in women. Test for linear trend was conducted by assigning the median value of exposure in each category to that category and treating the median value as a continuous variable in the regression model, with  $P < 0.05$  denoting a significant association.

Analyses were performed separately in each cohort first. We examined potential effect modifications by BMI, age, and aHEI score for both men and women, and menopausal status and postmenopausal hormone use for women. Meta-regressions were used to test for potential interactions, with  $P$  value  $< 0.05$  denoting effect modification. The tests for interaction were conducted in analyses within individual cohorts as well as in analyses based on pooled data from all three cohorts. Previous studies showed that coffee intake also contributes to total isoflavone intake<sup>20</sup> and was associated with a lower T2D risk in these cohorts.<sup>5</sup> To examine whether the association of isoflavones with diabetes risk may be due to coffee intake, we further calculated coffee-adjusted residuals of isoflavones using generalized equation estimation (GEE), and conducted a sensitivity analysis by using these residuals as the main exposure. All statistical tests were 2-sided and performed using SAS version 9.3 (SAS Institute Inc.). The meta-analysis was performed using STATA, version 9.2 (StataCorp).

## RESULTS

### Baseline characteristics according to soy food consumption

Baseline characteristics of the participants in each cohort according to soy food consumption are shown in Table 1. Most of the participants were non-consumers of soy foods at baseline in the three cohorts. Soy food consumers had a higher aHEI score, higher consumption of fruit, vegetables, and fish, lower consumption of meat and soda (including sugar-sweetened beverages), and were more physically active than non-consumers.

### The association of soy foods with risk of T2D

In the age-adjusted model, soy food consumption was inversely associated with risk of T2D. After multivariate adjustment, the association was attenuated and soy food consumption was non-significantly associated with a lower risk of T2D (Table 2). Compared with those who did not consume soy foods, the HR (95% CI) was 1.00 (0.93, 1.07) for those consuming < 1 serving/week of soy foods, and 0.93 (0.83, 1.03;  $P = 0.14$ ) for those consuming  $\geq 1$  serving/week of soy foods in the pooled analysis. We further examined the association separately with tofu and soy milk intake. Compared with non-consumers, the HR was 1.00 (95% CI: 0.93, 1.08) for those consuming < 1 serving/week of tofu, and 0.93 (95% CI: 0.84, 1.04) for those consuming  $\geq 1$  serving/week of tofu. For soy milk, compared with non-consumers, the HR was 0.92 (95% CI: 0.83, 1.02) for soy milk consumers. No significant associations of total soy foods, tofu, and soy milk with risk of T2D were found in the NHS, NHSII, and HPFS cohorts, and the associations did not vary significantly across the three cohorts (all  $P$  values for heterogeneity  $> 0.30$ ).

### **The association of isoflavone consumption with risk of T2D**

Total isoflavone consumption was significantly, inversely associated with risk of T2D (Table 3). As compared with the lowest quintile of isoflavones consumption, the HRs (95% CIs) were 0.89 (0.83, 0.96) for highest quintiles in the pooled analysis (P for trend = 0.009). We further evaluated individual isoflavones with meaningful intake levels in our cohorts (Table 4). For daidzein, the HR (95% CI) was 0.87 (0.81, 0.94) comparing extreme quintiles (P for trend = 0.0003). For genistein, the HR (95% CI) was 0.91 (0.85, 0.98) for the same comparison (P for trend = 0.02).

Of note, on average, regular soy food consumers ( $\geq 1$  serving/week) had higher isoflavone intake levels than participants in the highest quintile of total isoflavones, but the associations for soy food intake were not significant, suggesting that the association for isoflavones may not be linear at relatively high intake level. However, when we further examined the dose-response relationship between isoflavone intake and risk of T2D using spline regression, we did not observe a clear non-linear association (P for non-linearity: 0.76; P for trend = 0.02; Supplemental figure 1).

### **Stratified analysis**

We conducted analyses stratified by menopausal status (premenopausal vs postmenopausal; women only), BMI ( $< 30 \text{ kg/m}^2$  vs  $\geq 30 \text{ kg/m}^2$ ), age ( $< 60 \text{ y}$  vs  $\geq 60 \text{ y}$ ), and aHEI score ( $< \text{median}$  vs  $\geq \text{median}$ ), and no significant interactions were found between soy food and risk of T2D: P values for interaction were 0.20 for menopausal status, 0.78 for BMI, 0.34 for age, and 0.52 for aHEI score (Supplemental Table 1). No significant interactions were found between isoflavones consumption and these factors in

relation to T2D risk (Supplemental Table 2). We further tested effect modification by postmenopausal hormone use on the associations between consumptions of soy food and isoflavones and risk of T2D among postmenopausal women, but no significant effect modification was found. We performed further analyses restricted within postmenopausal women who were never users of hormone therapy and within women who were never users of soy supplements. The associations between intakes of soy foods and isoflavones and risk of T2D did not change substantially.

### **Sensitivity analysis**

As coffee was one of the food sources of isoflavones consumption,<sup>20</sup> we conducted sensitivity analysis using coffee-adjusted residuals of isoflavones consumption. Similar to the results of isoflavones, inverse associations of residual consumption of isoflavones, daidzein, and genistein with risk of T2D were found (Supplemental table 3). As soy food consumption was associated with a healthy lifestyle, we further repeated our analysis on the association of soy food and isoflavones with risk of T2D using propensity score analysis. The results did not change significantly comparing with the main analysis (Supplemental table 4, 5).

### **DISCUSSION**

In three large US cohorts of men and women, we found that isoflavone consumption was modestly associated with a lower risk of T2D, whereas the two major soy foods, i.e., tofu and soy milk, were not associated with T2D risk. These associations were independent of established and potential lifestyle and dietary risk factors of T2D.

The association of soy foods and isoflavones with risk of T2D has been investigated primarily among Asian populations who have much higher intake levels compared with Western populations, and the results have been largely mixed. Villegas et al. found that higher intakes of soybean and soy milk were significantly associated with a lower T2D incidence in the Shanghai Women's Health Study.<sup>21</sup> In the Singapore Chinese Health Study, Mueller et al. further documented that intakes of unsweetened soy products (servings/week), but not sugar-sweetened soy foods, were associated with a lower T2D risk.<sup>22</sup> However, a Japanese study found that intake of total soy foods (gram/day) with various soy protein densities was not associated with T2D risk,<sup>23</sup> whereas intake of total soy foods (gram/day) was associated with a higher T2D risk in a multi-ethnic population living in Hawaii.<sup>24</sup> Lastly, no association between total isoflavone intake and risk of T2D was found in the EPIC-InterAct Study.<sup>12</sup> The sources of heterogeneity in these findings are unclear, although study participant characteristics, different exclusion criteria, various cooking methods, and measurement error in soy food or isoflavone assessment may partially explain the mixed results. In the current analysis, we evaluated both major soy foods and isoflavones in relation to T2D risk and found the associations did not vary significantly across three cohorts of men and women.

In contrast to the paucity of evidence from long-term prospective observational studies, data from short-term clinical trials that examined the effects of soy foods or isoflavones on diabetes risk factors were abundant, and results were mixed. In a comprehensive meta-analysis of randomized controlled trials, supplementation of soy foods or isoflavones did not significantly lower fasting glucose or insulin levels, although a subgroup analysis showed that whole soy foods might reduce fasting glucose.<sup>9</sup> In

another meta-analysis that focused on premenopausal and postmenopausal non-Asian women who did not take hormone replacement therapy, isoflavone supplementation significantly lowered fasting insulin and HOMA-IR levels, although no effects were observed on fasting glucose levels.<sup>8</sup> In addition, soy products improved blood lipid profiles among diabetes patients, although the effects on glucose metabolism parameters were not substantiated.<sup>10</sup>

Isoflavones have a structure analogous to 17- $\beta$ -**estradiol**, which enables isoflavones bind to **estrogen** receptors  $\beta$  with  $10^3$ - $10^4$  less potency than estradiol.<sup>7</sup> Isoflavones exert either estrogenic or anti-estrogenic effect depending on the concentration of serum estradiol. Isoflavones exert estrogen-like effects when the concentration of endogenous estrogen is low, otherwise, isoflavones might have anti-estrogenic effect.<sup>25</sup> Isoflavones also activate nuclear receptors including peroxisome-proliferator activated receptors (PPAR)  $\alpha$  , PPAR  $\gamma$ , sterol regulated element binding protein, and liver X binding receptor to regulate lipid and glucose metabolism.<sup>26-28</sup> Isoflavones have been shown to improve hyperglycemia, glucose tolerance, and circulating insulin concentrations.<sup>29</sup> Isoflavones also stimulate the phosphorylation of AMP-activated protein kinase and acetyl-CoA carboxylase to increase glucose uptake and fatty acid oxidation.<sup>30</sup> Estrogen increases insulin sensitivity in the liver, promotes pancreas  $\beta$  cell proliferation and differentiation, modulates appetite and energy expenditure by regulating the expression of leptin and ghrelin, effects glucose disposal in muscle by upregulating expression of glucose transporter 4 and proteins involving the insulin signaling pathway,<sup>31</sup> and inhibits lipogenesis in adipose tissue by inhibiting the activity of lipoprotein lipase.<sup>32</sup> Whether isoflavones have those effects analogous to estrogen is speculated and needs further investigation.



Our study has several strengths. First, the analysis was based on three well-characterized large cohorts with detailed measurements of diet and lifestyle. Second, consumption of isoflavones and soy products was assessed every 4 years during the follow-up. The repeated measurements not only reduce measurement error but also represent long-term dietary habits. Third, the aHEI score was used to adjust for confounding of the overall diet quality. We also controlled for a wide range of lifestyle factors in the analysis. Our study also has several limitations. First, although we used the comprehensive USDA food composition database of isoflavones to derive isoflavone intake and included major soy foods in the current analysis, measurement error may still exist and may attenuate the true associations towards the null due to the longitudinal study design. Second, the low consumption levels of soy products in our cohorts (90% of participants were non-consumers), typically seen among Western populations, limited the statistical power for the analysis of soy food. Third, residual confounding by lifestyle factors (e.g., dietary factors, physical activity, and smoking) may still exist due to model-misspecification and measurement error of potential confounders. Last, our study was conducted primarily among white health professionals, and thus the results may not be generalizable to other populations.

In conclusion, our analysis showed that consumption of isoflavones, but not tofu or soy milk, was associated with a modest reduction in risk of T2D in three large cohorts of U.S. men and women. Further studies are needed to replicate these observations in other populations, especially those with similar isoflavone intake levels.

### **Contributors**

QS and FBH obtained funding from the National Institutes of Health. MD analyzed the data and wrote the first draft of the manuscript. AP, JEM, WCW, VM, BR, and EG contributed to the interpretation of the results and critical revision of the manuscript for important intellectual content and approved the final version of the manuscript. QS is the guarantor of this investigation.

### **Funding**

This study was funded by research grants CA186107, CA176726, CA167552, DK58845, DK58785, and DK082486 from the National Institutes of Health. Dr. Sun was supported by a career development award R00HL098459 from the National Heart, Lung, and Blood Institute. The funding sources had no role in study design; in the collection, analysis, and interpretation of data; in the writing of the report; or in the decision to submit the article for publication. The authors are not affiliated with the funding institutions.

### **Competing Interests**

None of the authors had any financial or personal conflict of interest to disclose.

### **Reference**

- 1 International Diabetes Federation. IDF Diabetes Atlas, 6th edn. *Brussels, Belgium: International Diabetes Federation* (2013).
- 2 Hu, F. B. Do functional foods have a role in the prevention of cardiovascular disease? *Circulation* **124**, 538-540, doi:10.1161/CIRCULATIONAHA.111.042721 (2011).
- 3 Muraki, I. *et al.* Fruit consumption and risk of type 2 diabetes: results from three prospective longitudinal cohort studies. *Bmj* **347**, f5001, doi:10.1136/bmj.f5001 (2013).
- 4 van Dam, R. M. & Hu, F. B. Coffee consumption and risk of type 2 diabetes: a systematic review. *JAMA : the journal of the American Medical Association* **294**, 97-104, doi:10.1001/jama.294.1.97 (2005).
- 5 Ding, M., Bhupathiraju, S. N., Chen, M., van Dam, R. M. & Hu, F. B. Caffeinated and decaffeinated coffee consumption and risk of type 2 diabetes: a systematic review and a dose-response meta-analysis. *Diabetes Care* **37**, 569-586, doi:10.2337/dc13-120337/2/569 [pii] (2014).

- 6 Bhagwat S, H. D., Holden JM, et al. USDA Database for the Isoflavone Content of Selected Foods Release 2.0. (2008).
- 7 Muthyala, R. S. *et al.* Equol, a natural estrogenic metabolite from soy isoflavones: convenient preparation and resolution of R- and S-equols and their differing binding and biological activity through estrogen receptors alpha and beta. *Bioorganic & medicinal chemistry* **12**, 1559-1567, doi:10.1016/j.bmc.2003.11.035 (2004).
- 8 Ricci, E., Cipriani, S., Chiaffarino, F., Malvezzi, M. & Parazzini, F. Effects of soy isoflavones and genistein on glucose metabolism in perimenopausal and postmenopausal non-Asian women: a meta-analysis of randomized controlled trials. *Menopause* **17**, 1080-1086, doi:10.1097/gme.0b013e3181dd05a9 (2010).
- 9 Liu, Z. M., Chen, Y. M. & Ho, S. C. Effects of soy intake on glycemic control: a meta-analysis of randomized controlled trials. *Am J Clin Nutr* **93**, 1092-1101, doi:10.3945/ajcn.110.007187 (2011).
- 10 Yang, B. *et al.* Systematic review and meta-analysis of soy products consumption in patients with type 2 diabetes mellitus. *Asia Pac J Clin Nutr* **20**, 593-602 (2011).
- 11 Sievenpiper, J. L. *et al.* Effect of non-oil-seed pulses on glycaemic control: a systematic review and meta-analysis of randomised controlled experimental trials in people with and without diabetes. *Diabetologia* **52**, 1479-1495, doi:10.1007/s00125-009-1395-7 (2009).
- 12 Zamora-Ros, R. *et al.* The association between dietary flavonoid and lignan intakes and incident type 2 diabetes in European populations: the EPIC-InterAct study. *Diabetes care* **36**, 3961-3970, doi:10.2337/dc13-0877 (2013).
- 13 Feskianich, D. *et al.* Reproducibility and validity of food intake measurements from a semiquantitative food frequency questionnaire. *J Am Diet Assoc* **93**, 790-796 (1993).
- 14 Rimm, E. B. *et al.* Reproducibility and validity of an expanded self-administered semiquantitative food frequency questionnaire among male health professionals. *American journal of epidemiology* **135**, 1114-1126; discussion 1127-1136 (1992).
- 15 Salvini, S. *et al.* Food-based validation of a dietary questionnaire: the effects of week-to-week variation in food consumption. *Int J Epidemiol* **18**, 858-867 (1989).
- 16 Willett, W. C. *et al.* Reproducibility and validity of a semiquantitative food frequency questionnaire. *American journal of epidemiology* **122**, 51-65 (1985).
- 17 Report of the Expert Committee on the Diagnosis and Classification of Diabetes Mellitus. *Diabetes care* **20**, 1183-1197 (1997).
- 18 Manson, J. E. *et al.* Physical activity and incidence of non-insulin-dependent diabetes mellitus in women. *Lancet* **338**, 774-778 (1991).
- 19 Hu, F. B. *et al.* Physical activity and television watching in relation to risk for type 2 diabetes mellitus in men. *Archives of internal medicine* **161**, 1542-1548 (2001).
- 20 Seema Bhagwat, D. B. H. a. J. M. H. USDA Database for the Isoflavone Content of Selected Foods Release 2.0 (2008).
- 21 Villegas, R. *et al.* Legume and soy food intake and the incidence of type 2 diabetes in the Shanghai Women's Health Study. *Am J Clin Nutr* **87**, 162-167 (2008).
- 22 Mueller, N. T. *et al.* Soy intake and risk of type 2 diabetes in Chinese Singaporeans [corrected]. *Eur J Nutr* **51**, 1033-1040, doi:10.1007/s00394-011-0276-2 (2012).
- 23 Nanri, A. *et al.* Soy product and isoflavone intakes are associated with a lower risk of type 2 diabetes in overweight Japanese women. *The Journal of nutrition* **140**, 580-586, doi:10.3945/jn.109.116020 (2010).

- 24 Morimoto, Y., Steinbrecher, A., Kolonel, L. N. & Maskarinec, G. Soy consumption is not protective against diabetes in Hawaii: the Multiethnic Cohort. *Eur J Clin Nutr* **65**, 279-282, doi:10.1038/ejcn.2010.228 (2011).
- 25 Messina, M. J. & Wood, C. E. Soy isoflavones, estrogen therapy, and breast cancer risk: analysis and commentary. *Nutrition journal* **7**, 17, doi:10.1186/1475-2891-7-17 (2008).
- 26 Mezei, O. *et al.* Soy isoflavones exert antidiabetic and hypolipidemic effects through the PPAR pathways in obese Zucker rats and murine RAW 264.7 cells. *The Journal of nutrition* **133**, 1238-1243 (2003).
- 27 Kim, S. *et al.* Hepatic gene expression profiles in a long-term high-fat diet-induced obesity mouse model. *Gene* **340**, 99-109, doi:10.1016/j.gene.2004.06.015 (2004).
- 28 Mezei, O., Li, Y., Mullen, E., Ross-Viola, J. S. & Shay, N. F. Dietary isoflavone supplementation modulates lipid metabolism via PPARalpha-dependent and - independent mechanisms. *Physiological genomics* **26**, 8-14, doi:10.1152/physiolgenomics.00155.2005 (2006).
- 29 Babu, P. V., Liu, D. & Gilbert, E. R. Recent advances in understanding the anti-diabetic actions of dietary flavonoids. *J Nutr Biochem* **24**, 1777-1789, doi:10.1016/j.jnutbio.2013.06.003 (2013).
- 30 Cederroth, C. R. *et al.* Dietary phytoestrogens activate AMP-activated protein kinase with improvement in lipid and glucose metabolism. *Diabetes* **57**, 1176-1185, doi:10.2337/db07-0630 (2008).
- 31 Barros, R. P. & Gustafsson, J. A. Estrogen receptors and the metabolic network. *Cell metabolism* **14**, 289-299, doi:10.1016/j.cmet.2011.08.005 (2011).
- 32 Misso, M. L. *et al.* Cellular and molecular characterization of the adipose phenotype of the aromatase-deficient mouse. *Endocrinology* **144**, 1474-1480, doi:10.1210/en.2002-221123 (2003).

Table 1. Baseline characteristics of participants by consumption of soy foods in the NHS, NHS II, and HPFS

	NHS (1998)			NHS II (1999)			HPFS (2002)		
	Non- consumer	< 1 serving/week	≥1 serving/ week	Non- consumer	< 1 serving/ week	≥1 serving/ week	Non- consumer	< 1 serving/ week	≥1 serving/ week
N	56,858	4,259	1,998	66,608	7,930	4,523	16,517	2,785	1,979
Age (year)	64	62	62	44	45	45	67	66	66
Total soy food (serving/d)	0	0.09	0.83	0	0.09	0.85	0	0.09	0.88
Soy milk (serving/d)	0	0	0.46	0	0.01	0.49	0	0.01	0.47
Tofu (serving/d)	0	0.08	0.37	0	0.08	0.36	0	0.08	0.41
Isoflavones (mg/d)	0.62	2.00	10.58	0.70	2.55	12.20	0.63	2.62	13.13
Daidzein (mg/d)	0.27	0.74	3.89	0.30	0.92	4.46	0.29	1.04	5.08
Genistein (mg/d)	0.31	1.10	5.05	0.37	1.41	5.87	0.31	1.25	5.70
Glycecin (mg/d)	0.04	0.17	1.72	0.04	0.21	1.88	0.03	0.33	2.34
Physical activity (MET-h/wk)	18	24	26	18	23	27	35	40	44
aHEI	53	61	64	50	59	63	56	63	68
Total energy intake (kcal/d)	1719	1816	1865	1809	1879	1928	1983	2009	2050
Fruits (serving/d)	2.36	2.93	3.12	1.76	2.26	2.57	2.57	3.09	3.52
Vegetables (serving/d)	3.07	3.92	4.25	3.19	4.10	4.60	3.36	4.04	4.48

Meat (serving/d)	1.24	1.01	0.85	1.42	1.17	0.92	1.58	1.33	1.12
Fish (serving/d)	0.22	0.29	0.30	0.19	0.27	0.28	0.28	0.38	0.38
Total soda (serving/d)	0.58	0.43	0.31	1.17	0.77	0.55	0.65	0.57	0.43
Coffee (cups/d)	1.82	1.62	1.30	1.51	1.56	1.31	1.64	1.45	1.22
Total alcoholic beverages (serving/d)	0.47	0.48	0.37	0.32	0.42	0.34	0.96	0.91	0.74
Dairy products (serving/d)	2.22	2.37	2.07	2.25	2.36	2.54	2.81	2.53	2.16
BMI (kg/m <sup>2</sup> )	26	25	25	24	23	22	26	26	25
Hypertension, %	41	37	35	13	11	8	44	43	39
Hypercholesterolemia, %	56	54	54	24	22	21	55	60	55
Family history of diabetes, %	27	27	25	35	34	32	21	22	21
Postmenopausal women, %	94	93	94	15	13	14	NA	NA	NA
Current menopausal hormone use, (% among total women)	53	54	47	14	11	10	NA	NA	NA
Current smokers, %	17	13	11	9	6	4	4	2	2
Race, Caucasian, %	98	92	91	97	93	92	97	91	92
Race, Asian, %	0	6	6	1	5	6	0	6	5

aHEI, Alternative Healthy Eating Index, with a higher score indicating healthier dietary pattern; BMI, body mass index; HPFS, Health Professionals Follow-up Study; MET, metabolic-equivalent task; NHS, Nurses' Health Study.

Table 2. Hazard ratios (HRs) for the associations between soy containing foods and risk of type 2 diabetes in the three cohorts

<b>Total soy food</b>	<b>Non-consumer</b>	<b>&lt; 1 serving/week</b>	<b>≥1 serving/week</b>	<b>P for trend</b>
<i>NHS (1998-2012)</i>				
Cases/Person-years	3,886/645,060	399/81,561	234/54,968	
Median intake (g/d)	0	0.05	0.43	
(range)	0	(0.02, 0.14)	(0.14, 6.00)	
Age-adjusted Model	1.00	0.86 (0.77, 0.95)	0.75 (0.66, 0.86)	0.004
Multivariate-adjusted Model	1.00	0.98 (0.88, 1.09)	0.97 (0.84, 1.11)	0.83
<i>NHS II (1999-2013)</i>				
Cases/Person-years	3,147/771,898	502/156,467	271/108,577	
Median intake (g/d)	0	0.07	0.43	
(range)	0	(0.02, 0.14)	(0.14, 8.50)	
Age-adjusted Model	1.00	0.79 (0.72, 0.87)	0.59 (0.52, 0.67)	< 0.001
Multivariate-adjusted Model	1.00	1.03 (0.93, 1.14)	0.92 (0.80, 1.05)	0.20
<i>HPFS (2002-2010)</i>				
Cases/Person-years	589/109,009	96/22,026	57/16,238	
Median intake (g/d)	0	0.07	0.50	
(range)	0	(0.04, 0.14)	(0.18, 10.50)	
Age-adjusted Model	1.00	0.83 (0.66, 1.02)	0.67 (0.51, 0.87)	0.003
Multivariate-adjusted Model	1.00	0.89 (0.71, 1.11)	0.88 (0.66, 1.17)	0.37
<i>Overall pooled</i>				
Multivariate-adjusted Model	1.00	1.00 (0.93, 1.07)	0.93 (0.83, 1.03)	0.14
<b>Tofu</b>	<b>Non-consumer</b>	<b>&lt; 1 serving/week</b>	<b>≥1 serving/week</b>	
<i>NHS (1998-2012)</i>				
Cases/Person-years	3,999/668,830	334/69,805	186/42,955	
Median intake (g/d)	0	0.05	0.22	
(range)	0	(0.02, 0.07)	(0.09, 6.00)	
Age-adjusted Model	1.00	0.84 (0.75, 0.94)	0.75 (0.65, 0.87)	0.001
Multivariate-adjusted Model	1.00	0.98 (0.88, 1.10)	0.98 (0.84, 1.14)	0.97
<i>NHS II (1999-2013)</i>				
Cases/Person-years	3,282/812,630	416/130,915	222/93,398	
Median intake (g/d)	0	0.05	0.22	

(range)	0	(0.02, 0.07)	(0.09, 6.00)	
Age-adjusted Model	1.00	0.79 (0.71, 0.87)	0.58 (0.50, 0.66)	< 0.001
Multivariate-adjusted Model	1.00	1.02 (0.92, 1.14)	0.91 (0.78, 1.05)	0.25
<i>HPFS (2002-2010)</i>				
Cases/Person-years	614/114,355	77/17,834	51/15,083	
Median intake (g/d)	0	0.07	0.29	
(range)	0	(0.04, 0.07)	(0.11, 6.00)	
Age-adjusted Model	1.00	0.83 (0.65, 1.05)	0.64 (0.48, 0.85)	0.002
Multivariate-adjusted Model	1.00	0.90 (0.71, 1.15)	0.81 (0.60, 1.10)	0.23
<i>Overall pooled</i>				
Multivariate-adjusted Model	1.00	1.00 (0.93, 1.08)	0.93 (0.84, 1.04)	0.19
<b>Soy milk</b>	<b>Non-consumer</b>	<b>Consumer</b>		
<i>NHS (1998-2012)</i>				
Cases/Person-years	4,287/726,315	232/55,274		
Median intake (g/d)	0	0.22		
(range)	0	(0.02, 6.00)		
Age-adjusted Model	1.00	0.79 (0.69, 0.90)		0.24
Multivariate-adjusted Model	1.00	0.96 (0.84, 1.10)		0.84
<i>NHS II (1999-2013)</i>				
Cases/Person-years	3,621/920,118	299/116,823		
Median intake (g/d)	0	0.17		
(range)	0	(0.02, 6.00)		
Age-adjusted Model	1.00	0.65 (0.58, 0.73)		< 0.001
Multivariate-adjusted Model	1.00	0.90 (0.80, 1.02)		0.11
<i>HPFS (2002-2010)</i>				
Cases/Person-years	691/133,130	51/14,143		
Median intake (g/d)	0	0.43		
(range)	0	(0.04, 6.00)		
Age-adjusted Model	1.00	0.72 (0.54, 0.95)		0.02
Multivariate-adjusted Model	1.00	0.93 (0.69, 1.24)		0.58
<i>Overall pooled</i>				
Multivariate-adjusted Model	1.00	0.92 (0.83, 1.02)		0.11

Abbreviations: NHS, Nurses' Health Study; HPFS, Health Professionals Follow-up Study;



Multivariate-adjusted model: adjusted for race (Caucasians, African Americans, Asian Americans, and others), family history of T2D (yes vs. no), baseline disease status (hypertension, hypercholesterolemia), body mass index (<21, 21-22.9, 23-24.9, 25-29.9, 30-34.9,  $\geq 35$  kg/m<sup>2</sup>), physical activity (quintiles), overall dietary pattern (Alternative Healthy Eating Index score, in quintiles), total energy intake (quintiles), coffee consumption (quintiles), smoking status (never, former, current 1-14 cigarettes/d, current >14 cigarettes/d). Menopausal status (yes vs. no) and postmenopausal hormone use (yes vs. no) were further adjusted for in women.

Table 3. Associations between isoflavone consumption and risk of type 2 diabetes in the three cohorts

	Q1	Q2	Q3	Q4	Q5	P for trend
<b>NHS (1998-2012)</b>						
Cases/Person-years	894/151,060	1042/160,232	934/157,173	848/155,786	801/157,339	
Median intake	0.17	0.29	0.40	0.62	2.78	
(range #) (mg/d)	(0.01, 0.44)	(0.17, 0.59)	(0.26, 0.80)	(0.37, 1.78)	(0.57, 76.57)	
Age-adjusted Model	1.00	1.09 (1.00, 1.19)	1.00 (0.91, 1.10)	0.92 (0.83, 1.01)	0.85 (0.77, 0.93)	< 0.001
Multivariate-adjusted Model	1.00	1.08 (0.99, 1.18)	1.02 (0.93, 1.12)	0.96 (0.87, 1.05)	0.97 (0.88, 1.07)	0.13
<b>NHS II (1999-2013)</b>						
Cases/Person-years	913/199,825	905/210,571	753/209,716	724/208,238	625/208,591	
Median intake (mg/d)	0.17	0.31	0.48	1.10	5.73	
(range)	(0.01, 0.46)	(0.17, 0.75)	(0.27, 1.50)	(0.42, 3.97)	(1.14, 130.50)	
Age-adjusted Model	1.00	0.94 (0.86, 1.03)	0.77 (0.70, 0.85)	0.74 (0.67, 0.81)	0.63 (0.57, 0.70)	< 0.001
Multivariate-adjusted Model	1.00	0.95 (0.86, 1.04)	0.82 (0.74, 0.90)	0.85 (0.77, 0.94)	0.85 (0.76, 0.95)	0.11
<b>HPFS (2002-2010)</b>						
Cases/Person-years	170/29,598	144/29,356	151/29,252	160/29,530	117/29,537	
Median intake	0.31	0.47	0.64	1.10	5.09	
(range) (mg/d)	(0.01, 0.49)	(0.35, 0.66)	(0.48, 0.93)	(0.66, 2.27)	(1.87, 238.02)	
Age-adjusted Model	1.00	0.86 (0.68, 1.07)	0.90 (0.73, 1.13)	0.95 (0.76, 1.18)	0.69 (0.54, 0.87)	0.004
Multivariate-adjusted Model	1.00	0.83 (0.67, 1.04)	0.87 (0.70, 1.09)	0.93 (0.74, 1.15)	0.80 (0.62, 1.02)	0.24
<b>Pooled</b>						
Multivariate-adjusted Model	1.00	0.99 (0.93, 1.06)	0.91 (0.86, 0.97)	0.91 (0.85, 0.97)	0.89 (0.83, 0.96)	0.009

#: Overlap of range was due to that the quintile was divided within each time interval of the Cox model.

NHS, Nurses' Health Study; HPFS, Health Professionals Follow-up Study

Multivariate-adjusted for race (Caucasians, African Americans, Asian Americans, and others), family history of T2D (yes vs. no), baseline disease status (hypertension, hypercholesterolemia), BMI (<21, 21-22.9, 23-24.9, 25-29.9, 30-34.9, ≥35 kg/m<sup>2</sup>), physical activity (quintiles), overall dietary pattern (AHEI score, in quintiles), total energy intake (quintiles), smoking status (never, former,

current (1-14 cigarettes/d), current (>14 cigarettes/d)). Menopausal status (yes vs. no) and postmenopausal hormone use (yes vs. no) were further adjusted for in women.

Table 4. Hazard ratio (HR) for the association between subtypes of isoflavone consumption and risk of type 2 diabetes in the three cohorts

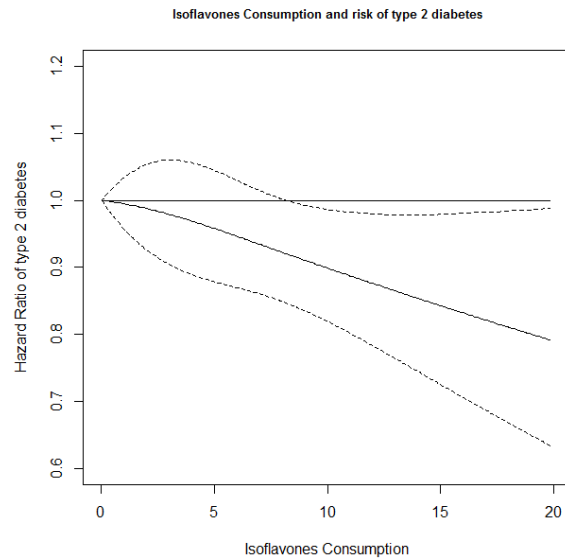
	Q1	Q2	Q3	Q4	Q5	P for trend
<b>Daidzein</b>						
<i>NHS (1998-2012)</i>	904/148,554	1069/163,641	947/155,285	811/156,098	788/158,012	
Median intake	0.08	0.15	0.23	0.33	1.05	
(range #) (mg/d)	(0.01, 0.20)	(0.08, 0.28)	(0.14, 0.39)	(0.22, 0.72)	(0.35, 25.46)	
Age-adjusted Model	1.00	1.06 (0.97, 1.15)	1.00 (0.91, 1.09)	0.85 (0.77, 0.93)	0.80 (0.73, 0.88)	< 0.001
Multivariate-adjusted Model	1.00	1.05 (0.96, 1.15)	1.02 (0.93, 1.12)	0.90 (0.81, 0.99)	0.92 (0.84, 1.02)	0.01
<i>NHS II (1999-2013)</i>	908/198,618	966/209,843	727/210,988	699/208,319	620/209,174	
Median intake (mg/d)	0.07	0.16	0.27	0.49	2.09	
(range)	(0.01, 0.22)	(0.08, 0.36)	(0.14, 0.63)	(0.25, 1.50)	(0.50, 41.70)	
Age-adjusted Model	1.00	1.00 (0.92, 1.10)	0.74 (0.67, 0.82)	0.71 (0.64, 0.78)	0.62 (0.56, 0.69)	< 0.001
Multivariate-adjusted Model	1.00	0.99 (0.91, 1.09)	0.80 (0.73, 0.89)	0.84 (0.75, 0.93)	0.85 (0.76, 0.94)	0.03
<i>HPFS (2002-2010)</i>	166/29,178	158/30,164	153/28,814	146/29,621	119/29,496	
Median intake	0.14	0.22	0.32	0.53	1.98	
(range) (mg/d)	(0.01, 0.22)	(0.16, 0.31)	(0.24, 0.45)	(0.35, 0.92)	(0.77, 75.11)	
Age-adjusted Model	1.00	0.91 (0.74, 1.14)	0.94 (0.76, 1.17)	0.87 (0.69, 1.08)	0.71 (0.56, 0.89)	0.005
Multivariate-adjusted Model	1.00	0.90 (0.72, 1.12)	0.91 (0.73, 1.14)	0.83 (0.67, 1.04)	0.81 (0.64, 1.04)	0.23
<i>Pooled</i>						
Multivariate-adjusted Model	1.00	1.01 (0.95, 1.07)	0.91 (0.85, 0.97)	0.86 (0.81, 0.92)	0.87 (0.81, 0.94)	0.0003
<b>Genistein</b>						
<i>NHS (1998-2012)</i>	910/156,013	965/158,628	913/153,330	927/156,833	804/156,785	
Median intake (mg/d)	0.08	0.13	0.17	0.26	1.38	
(range)	(0.01, 0.21)	(0.09, 0.27)	(0.12, 0.37)	(0.15, 0.87)	(0.22, 44.72)	
Age-adjusted Model	1.00	1.04 (0.95, 1.14)	1.03 (0.94, 1.13)	1.01 (0.92, 1.11)	0.87 (0.79, 0.96)	0.003
Multivariate-adjusted Model	1.00	1.04 (0.95, 1.14)	1.03 (0.94, 1.13)	1.02 (0.93, 1.12)	0.98 (0.89, 1.08)	0.27
<i>NHS II (1999-2013)</i>	903/205,677	820/196,157	816/220,327	759/206,270	622/208,510	
Median intake (mg/d)	0.09	0.15	0.19	0.54	2.87	
(range)	(0.01, 0.22)	(0.09, 0.34)	(0.12, 0.74)	(0.17, 1.99)	(0.58, 78.62)	
Age-adjusted Model	1.00	0.95 (0.87, 1.05)	0.84 (0.76, 0.92)	0.82 (0.74, 0.90)	0.66 (0.59, 0.73)	< 0.001

Multivariate-adjusted Model	1.00	0.97 (0.88, 1.07)	0.86 (0.78, 0.95)	0.91 (0.82, 1.00)	0.87 (0.78, 0.97)	0.11
<i>HPFS (2002-2010)</i>	158/28,674	148/29,339	161/29,819	158/29,945	117/29,496	
Median intake	0.15	0.21	0.28	0.51	2.29	
(range) (mg/d)	(0.01, 0.24)	(0.17, 0.32)	(0.22, 0.43)	(0.29, 1.08)	(0.88, 142.09)	
Age-adjusted Model	1.00	0.91 (0.73, 1.14)	0.97 (0.78, 1.22)	0.96 (0.77, 1.20)	0.71 (0.56, 0.91)	0.005
Multivariate-adjusted Model	1.00	0.90 (0.72, 1.13)	0.94 (0.75, 1.17)	0.96 (0.77, 1.21)	0.84 (0.65, 1.07)	0.28
<i>Pooled</i>						
Multivariate-adjusted Model	1.00	0.99 (0.93, 1.06)	0.95 (0.89, 1.01)	0.97 (0.91, 1.03)	0.91 (0.85, 0.98)	0.02

#: Overlap of range was due to that the quintile was divided within each time interval of the Cox model.

Age-adjusted Model: age-adjusted model.

Multivariate-adjusted model: multivariate model adjusted for race (Caucasians, African Americans, Asian Americans, and others), family history of T2D (yes vs. no), baseline disease status (hypertension, hypercholesterolemia), BMI (<21, 21-22.9, 23-24.9, 25-29.9, 30-34.9,  $\geq 35$  kg/m<sup>2</sup>), physical activity (quintiles), overall dietary pattern (AHEI score, in quintiles), total energy intake (quintiles), smoking status (never, former, current (1-14 cigarettes/d), current (>14 cigarettes/d)). Menopausal status (yes vs. no) and postmenopausal hormone use (yes vs. no) were further adjusted for in women.



Supplemental Figure 1. The association between isoflavones consumption and risk of type 2 diabetes (T2D) in a dose response manner by pooling the three cohorts.

Multivariate-adjusted model: multivariate model adjusted for race (Caucasians, African Americans, Asian Americans, and others), family history of T2D (yes vs. no), baseline disease status (hypertension, hypercholesterolemia), BMI (<21, 21-22.9, 23-24.9, 25-29.9, 30-34.9,  $\geq 35$  kg/m<sup>2</sup>), physical activity (quintiles), overall dietary pattern (AHEI score, in quintiles), total energy intake (quintiles), smoking status (never, former, current (1-14 cigarettes/d), current (>14 cigarettes/d)). Menopausal status (yes vs. no), and postmenopausal hormone use (yes vs. no) were further adjusted for in women.

Supplemental Table 1. Stratified analysis of the association between total soy food consumption (consumer vs. non-consumer) and risk of type 2 diabetes (T2D).

	NHS	NHS II	HPFS	Pooled	P for interaction
<b>Menopausal status</b>					
Premenopausal	1.09 (0.95, 1.26)	1.06 (0.93, 1.20)		1.07 (0.98, 1.18)	0.20
Postmenopausal	0.94 (0.84, 1.05)	0.94 (0.83, 1.05)		0.94 (0.87, 1.02)	
<b>Postmenopausal hormone and isoflavones supplement use</b>					
Neither	1.08 (0.94, 1.25)	0.99 (0.89, 1.11)		1.02 (0.94, 1.12)	0.78
Only postmenopausal hormone	0.87 (0.75, 1.00)	1.40 (0.76, 2.55)		1.00 (0.66, 1.54)	
Only isoflavones supplement	1.05 (0.86, 1.29)	1.03 (0.90, 1.19)		1.04 (0.92, 1.16)	
Both	2.20 (0.71, 6.84)	1.11 (0.55, 2.21)		1.34 (0.74, 2.45)	
<b>BMI</b>					
BMI (< 30 kg/m <sup>2</sup> )	1.00 (0.87, 1.14)	0.82 (0.68, 0.98)	0.96 (0.76, 1.20)	0.94 (0.85, 1.03)	0.78
BMI (≥ 30 kg/m <sup>2</sup> )	0.97 (0.85, 1.11)	1.07 (0.96, 1.19)	0.66 (0.46, 0.94)	1.01 (0.93, 1.09)	
<b>Age</b>					
Age (< 60 y)	0.94 (0.53, 1.66)	1.02 (0.92, 1.12)	0.95 (0.72, 1.25)	1.01 (0.92, 1.11)	0.34
Age (≥ 60 y)	0.98 (0.89, 1.07)	0.91 (0.77, 1.09)	0.82 (0.63, 1.05)	0.95 (0.88, 1.03)	
<b>aHEI score</b>					
aHEI (< median)	0.92 (0.78, 1.09)	0.96 (0.85, 1.10)	0.66 (0.43, 1.01)	0.93 (0.84, 1.02)	0.52
aHEI (≥ median)	0.96 (0.87, 1.07)	0.98 (0.87, 1.09)	0.94 (0.76, 1.15)	0.97 (0.90, 1.04)	

Multivariate-adjusted model : multivariate model adjusted for race (Caucasians, African Americans, Asian Americans, and others), family history of T2D (yes vs. no), baseline disease status (hypertension, hypercholesterolemia), BMI (<21, 21-22.9, 23-24.9, 25-29.9, 30-34.9, ≥35 kg/m<sup>2</sup>), physical activity (quintiles), overall dietary pattern (AHEI score, in quintiles), total energy intake (quintiles), smoking status (never, former, current (1-14 cigarettes/d), current (>14 cigarettes/d)). Menopausal status (yes vs. no), and postmenopausal hormone use (yes vs. no) were further adjusted for in women.

Supplemental Table 2. Stratified analysis of the association between consumption of isoflavones (dichotomized) and risk of type 2 diabetes (T2D)

	NHS	NHSII	HPFS	Pooled	P for interaction
<b>Menopausal status</b>					
Premenopausal	1.07 (0.93, 1.23)	0.85 (0.78, 0.94)		0.91 (0.84, 0.98)	0.56
Postmenopausal	0.89 (0.83, 0.96)	0.87 (0.78, 0.96)		0.88 (0.83, 0.93)	
<b>Postmenopausal hormone and isoflavones supplement use</b>					
Neither	1.07 (0.93, 1.24)	0.85 (0.78, 0.93)		0.95 (0.76, 1.19)	0.89
Only postmenopausal hormone	NA	0.89 (0.48, 1.63)		0.89 (0.48, 1.64)	
Only isoflavones supplement	0.87 (0.79, 0.95)	0.91 (0.80, 1.04)		0.88 (0.82, 0.95)	
Both	3.41 (1.03, 11.31)	1.27 (0.60, 2.67)		1.85 (0.72, 4.75)	
<b>BMI</b>					
BMI (< 30 kg/m <sup>2</sup> )	0.90 (0.82, 0.99)	0.81 (0.69, 0.95)	1.05 (0.80, 1.39)	0.89 (0.82, 0.96)	0.67
BMI (≥ 30 kg/m <sup>2</sup> )	0.96 (0.87, 1.05)	0.89 (0.82, 0.97)	0.74 (0.51, 1.07)	0.91 (0.86, 0.97)	
<b>Age</b>					
Age (< 60 y)	0.91 (0.66, 1.24)	0.88 (0.82, 0.95)	0.82 (0.60, 1.11)	0.88 (0.82, 0.94)	0.67
Age (≥ 60 y)	0.93 (0.87, 0.99)	0.78 (0.66, 0.92)	1.17 (0.86, 1.59)	0.92 (0.87, 0.97)	
<b>aHEI score</b>					
aHEI (< median)	0.90 (0.83, 0.99)	0.87 (0.80, 0.95)	0.94 (0.70, 1.26)	0.90 (0.84, 0.95)	0.69
aHEI (≥ median)	0.95 (0.87, 1.04)	0.83 (0.73, 0.94)	1.05 (0.76, 1.44)	0.91 (0.85, 0.99)	

Multivariate-adjusted model : multivariate model adjusted for race (Caucasians, African Americans, Asian Americans, and others), family history of T2D (yes vs. no), baseline disease status (hypertension, hypercholesterolemia), BMI (<21, 21-22.9, 23-24.9, 25-29.9, 30-34.9, ≥35 kg/m<sup>2</sup>), physical activity (quintiles), overall dietary pattern (AHEI score, in quintiles), total energy intake (quintiles), smoking status (never, former, current (1-14 cigarettes/d), current (>14 cigarettes/d)). Menopausal status (yes vs. no), and postmenopausal hormone use (yes vs. no) were further adjusted for in women.



Supplemental Table 3. Hazard ratio (HR) for the associations between residual isoflavones consumption after adjusting for coffee and risk of type 2 diabetes in the three cohorts

	Q1	Q2	Q3	Q4	Q5	P for trend
<b>Isoflavones</b>						
<i>NHS (1998-2012)</i>						
Multivariate-adjusted Model	1.00	1.06 (0.97, 1.16)	1.00 (0.91, 1.09)	0.89 (0.81, 0.98)	0.92 (0.83, 1.01)	0.14
<i>NHS II (1999-2013)</i>						
Multivariate-adjusted Model	1.00	0.94 (0.86, 1.03)	0.81 (0.73, 0.89)	0.85 (0.77, 0.94)	0.84 (0.76, 0.94)	0.11
<i>HPFS (2002-2010)</i>						
Multivariate-adjusted Model	1.00	0.82 (0.65, 1.03)	0.85 (0.68, 1.06)	0.80 (0.63, 1.01)	0.76 (0.59, 0.98)	0.21
<i>Pooled</i>						
Multivariate-adjusted Model	1.00	0.97 (0.86, 1.09)	0.89 (0.76, 1.04)	0.86 (0.81, 0.92)	0.87 (0.80, 0.94)	0.03
<b>Daidzein</b>						
<i>NHS (1998-2012)</i>						
Multivariate-adjusted Model	1.00	1.05 (0.96, 1.14)	1.04 (0.95, 1.14)	0.90 (0.82, 0.99)	0.93 (0.84, 1.02)	0.01
<i>NHS II (1999-2013)</i>						
Multivariate-adjusted Model	1.00	1.00 (0.91, 1.10)	0.90 (0.81, 0.99)	0.92 (0.83, 1.02)	0.89 (0.79, 0.99)	0.03
<i>HPFS (2002-2010)</i>						
Multivariate-adjusted Model	1.00	0.82 (0.66, 1.03)	0.89 (0.71, 1.10)	0.71 (0.57, 0.90)	0.73 (0.57, 0.94)	0.17
<i>Pooled</i>						
Multivariate-adjusted Model	1.00	0.99 (0.89, 1.10)	0.95 (0.85, 1.07)	0.87 (0.78, 0.97)	0.89 (0.81, 0.98)	0.001
<b>Genistein</b>						
<i>NHS (1998-2012)</i>						
Multivariate-adjusted	1.00	1.08 (0.99, 1.18)	0.95 (0.87, 1.05)	0.92 (0.84, 1.01)	0.91 (0.83, 1.01)	0.27

Model						
<i>NHS II (1999-2013)</i>						
Multivariate-adjusted Model	1.00	0.95 (0.86, 1.04)	0.79 (0.71, 0.87)	0.84 (0.76, 0.93)	0.83 (0.74, 0.92)	0.11
<i>HPFS (2002-2010)</i>						
Multivariate-adjusted Model	1.00	0.84 (0.67, 1.05)	0.87 (0.69, 1.09)	0.84 (0.67, 1.05)	0.77 (0.60, 0.99)	0.21
<i>Pooled</i>						
Multivariate-adjusted Model	1.00	0.98 (0.87, 1.11)	0.87 (0.76, 0.99)	0.88 (0.82, 0.94)	0.86 (0.79, 0.94)	0.04

Multivariate-adjusted model: multivariate model adjusted for race (Caucasians, African Americans, Asian Americans, and others), family history of T2D (yes vs. no), baseline disease status (hypertension, hypercholesterolemia), BMI (<21, 21-22.9, 23-24.9, 25-29.9, 30-34.9,  $\geq 35$  kg/m<sup>2</sup>), physical activity (quintiles), overall dietary pattern (AHEI score, in quintiles), total energy intake (quintiles), smoking status (never, former, current (1-14 cigarettes/d), current (>14 cigarettes/d)). Menopausal status (yes vs. no) and postmenopausal hormone use (yes vs. no) were further adjusted for in women.

Supplemental Table 4. Hazard ratios (HRs) for the associations between soy containing foods and risk of type 2 diabetes in the three cohorts using propensity score analysis

<b>Total soy food</b>	<b>Non-consumer</b>	<b>&lt; 1 serving/week</b>	<b>≥1 serving/week</b>	<b>P for trend</b>
NHS (1998-2012)	1.00	0.97 (0.87, 1.08)	0.95 (0.83, 1.09)	0.36
NHS II (1999-2013)	1.00	1.04 (0.94, 1.15)	0.87 (0.77, 1.00)	0.05
HPFS (2002-2010)	1.00	0.98 (0.79, 1.23)	0.88 (0.66, 1.17)	0.38
Overall pooled	1.00	1.01 (0.95, 1.09)	0.90 (0.83, 0.99)	0.03
<b>Tofu</b>	<b>Non-consumer</b>	<b>&lt; 1 serving/week</b>	<b>≥1 serving/week</b>	
NHS (1998-2012)	1.00	0.97 (0.87, 1.09)	0.95 (0.81, 1.11)	0.58
NHS II (1999-2013)	1.00	1.03 (0.93, 1.14)	0.87 (0.75, 1.00)	0.09
HPFS (2002-2010)	1.00	0.97 (0.76, 1.23)	0.83 (0.61, 1.11)	0.22
Overall pooled	1.00	1.01 (0.94, 1.09)	0.91 (0.82, 1.00)	0.05
<b>Soy milk</b>	<b>Non-consumer</b>	<b>Consumer</b>		
NHS (1998-2012)	1.00	0.95 (0.83, 1.09)		0.91
NHS II (1999-2013)	1.00	0.89 (0.79, 1.01)		0.06
HPFS (2002-2010)	1.00	0.92 (0.69, 1.24)		0.56
Overall pooled	1.00	0.91 (0.84, 1.00)		0.11

Abbreviations: NHS, Nurses' Health Study; HPFS, Health Professionals Follow-up Study;

Covariates adjusted for in the propensity score model: adjusted for race (Caucasians, African Americans, Asian Americans, and others), family history of T2D (yes vs. no), baseline disease status (hypertension, hypercholesterolemia), body mass index (<21, 21-22.9, 23-24.9, 25-29.9, 30-34.9, ≥35 kg/m<sup>2</sup>), physical activity (quintiles), overall dietary pattern (Alternative Healthy Eating Index score, in quintiles), total energy intake (quintiles), coffee consumption (quintiles), smoking status (never, former, current 1-14 cigarettes/d, current >14 cigarettes/d). Menopausal status (yes vs. no) and postmenopausal hormone use (yes vs. no) were further adjusted for in women.

Supplemental Table 5. Associations between isoflavone consumption and risk of type 2 diabetes in the three cohorts using propensity score analysis

	Q1	Q2	Q3	Q4	Q5	P for trend
<b>Isoflavones</b>						
NHS (1998-2012)	1.00	1.07 (0.97, 1.17)	1.01 (0.92, 1.10)	0.94 (0.85, 1.03)	0.95 (0.86, 1.05)	0.06
NHS II (1999-2013)	1.00	0.98 (0.90, 1.08)	0.83 (0.75, 0.91)	0.87 (0.79, 0.96)	0.89 (0.80, 0.99)	0.29
HPFS (2002-2010)	1.00	0.85 (0.68, 1.06)	0.88 (0.71, 1.10)	0.97 (0.78, 1.21)	0.86 (0.67, 1.10)	0.49
Pooled	1.00	0.99 (0.89, 1.10)	0.89 (0.80, 0.99)	0.89 (0.84, 0.95)	0.91 (0.85, 0.97)	0.06
<b>Genistein</b>						
NHS (1998-2012)	1.00	1.03 (0.94, 1.12)	1.01 (0.92, 1.10)	1.00 (0.91, 1.10)	0.95 (0.87, 1.05)	0.15
NHS II (1999-2013)	1.00	0.98 (0.89, 1.08)	0.89 (0.81, 0.98)	0.95 (0.86, 1.04)	0.90 (0.80, 1.00)	0.16
HPFS (2002-2010)	1.00	0.92 (0.73, 1.15)	0.99 (0.80, 1.24)	1.02 (0.82, 1.27)	0.91 (0.71, 1.17)	0.50
Pooled	1.00	1.00 (0.94, 1.07)	0.96 (0.87, 1.06)	0.99 (0.93, 1.05)	0.93 (0.87, 1.00)	0.03
<b>Daidzein</b>						
NHS (1998-2012)	1.00	1.04 (0.95, 1.14)	1.01 (0.92, 1.11)	0.88 (0.80, 0.97)	0.91 (0.83, 1.00)	0.01
NHS II (1999-2013)	1.00	1.08 (0.98, 1.18)	0.84 (0.76, 0.93)	0.89 (0.81, 0.99)	0.92 (0.83, 1.03)	0.20
HPFS (2002-2010)	1.00	0.91 (0.73, 1.13)	0.91 (0.73, 1.13)	0.90 (0.72, 1.12)	0.89 (0.70, 1.14)	0.55
Pooled	1.00	1.05 (0.99, 1.11)	0.90 (0.82, 0.99)	0.88 (0.82, 0.94)	0.91 (0.85, 0.98)	0.03

NHS, Nurses' Health Study; HPFS, Health Professionals Follow-up Study

Covariates adjusted for in the propensity score model: race (Caucasians, African Americans, Asian Americans, and others), family history of T2D (yes vs. no), baseline disease status (hypertension, hypercholesterolemia), BMI (<21, 21-22.9, 23-24.9, 25-29.9, 30-34.9,  $\geq 35$  kg/m<sup>2</sup>), physical activity (quintiles), overall dietary pattern (AHEI score, in quintiles), total energy intake (quintiles), smoking status (never, former, current (1-14 cigarettes/d), current (>14 cigarettes/d)). Menopausal status (yes vs. no) and postmenopausal hormone use (yes vs. no) were further adjusted for in women.

## CHAPTER 4

---

### URINARY ISOFLAVONES AND RISK OF TYPE 2 DIABETES: A PROSPECTIVE INVESTIGATION IN U.S. WOMEN

---

Ming Ding <sup>1</sup>, Adrian A. Franke <sup>2</sup>, Bernard A. Rosner <sup>3,4</sup>, Edward Giovannucci <sup>1,3</sup>, Rob M. van Dam <sup>1,5</sup>, Shelley S. Tworoger <sup>3,6</sup>, Frank B. Hu <sup>1,3,6</sup>, Qi Sun <sup>1,3</sup>

<sup>1</sup>Department of Nutrition, Harvard School of Public Health, Boston, MA

<sup>2</sup>University of Hawai‘i Cancer Center, Honolulu, HI, United States

<sup>3</sup>Channing Division of Network Medicine, Brigham and Women’s Hospital and Harvard Medical School, Boston, MA

<sup>4</sup>Department of Statistics, Harvard School of Public Health, Boston, MA

<sup>5</sup>Saw Swee Hock School of Public Health and Yong Loo Lin School of Medicine, National University of Singapore and National University Health System, Singapore

<sup>6</sup>Department of Epidemiology, Harvard School of Public Health, Boston, MA

## ABSTRACT

**Background** To examine the association between urinary excretion of isoflavonoids and risk of type 2 diabetes (T2D),

**Method** we conducted a nested case-control study among 1,111 T2D pairs identified during 1995 - 2008 in the Nurses' Health Study (NHS) and NHSII who were free of diabetes, cardiovascular disease, and cancer at urine sample collection. Urinary excretion of daidzein and genistein, as well as their metabolites desmethylangolensin, dihydrogenistein (DHGE), dihydrodaidzein (DHDE) was assayed by liquid chromatography mass spectrometry. Incident self-reported T2D cases were confirmed using a validated questionnaire.

**Results** Higher urinary excretion of daidzein and genistein was associated with a lower risk of T2D in the combined cohorts. Comparing extreme tertiles of the urinary markers, the odds ratios (ORs) of T2D were 0.71 (95% CI: 0.55, 0.93) for daidzein and 0.74 (95% CI: 0.56, 0.97) for genistein, although the test for linear trend was not significant for genistein ( $P$  for trend = 0.03 and 0.15, respectively). DMA, DHDE, and DHGE were not significantly associated with risk of T2D. The inverse association of daidzein with T2D risk was stronger among postmenopausal women who did not use hormone replacement therapy ( $P$  for interaction = 0.001): the OR (95% CI) was 0.58 (0.34, 0.97;  $P$  for trend = 0.06) comparing extreme tertiles among these women.

**Conclusion** In conclusion, urinary excretion of daidzein might be associated with lower T2D risk, especially among postmenopausal women who did not use hormone. However, further research is warranted to replicate these observations.

## INTRODUCTION

Type 2 diabetes (T2D) is a chronic disease with increasing prevalence worldwide. The total number of diabetes patients is estimated to reach 592 million globally by the year 2035<sup>(1)</sup>. Seeking effective dietary and lifestyle measures for T2D prevention has been a priority to counteract the increasing diabetes prevalence and incidence<sup>(36; 8; 23)</sup>. Women may be at a particularly high risk at middle-life when menopause and aging may jointly increase the risk of T2D<sup>(5; 35; 9)</sup>. Large clinical trials and prospective cohort studies have consistently shown that hormone replacement therapy may reduce risk of T2D in postmenopausal women<sup>(10; 14 15 31)</sup>. However, it is unclear whether natural phytoestrogens, such as isoflavones, may be associated with T2D risk.

Clinical trials showed that isoflavone supplements did not improve glucose control<sup>(12 39)</sup>, although these clinical trials were limited by small sample size and short duration of follow-up. Several cohort studies have evaluated the associations of isoflavone intakes assessed using food frequency questionnaire (FFQ) with risk of T2D, and mixed results were observed<sup>(26; 24; 40)</sup>. One potential reason for the inconsistent findings in the observational studies may lie in the difficulties of using FFQs to assess isoflavone intakes. Soy foods are the main source of isoflavone intakes while other foods contain various amount of isoflavones as well<sup>(3)</sup>. Isoflavone intake estimated from FFQs is thus subject to measurement errors<sup>(33)</sup>, especially in Western populations who infrequently consume soy foods. In addition, isoflavone intake estimated from FFQs does not take into account inter-individual variations in bioavailability<sup>(41)</sup>. In addition, FFQs cannot be used to estimate the gut microbiota metabolites of isoflavones, including O-desmethylangolensin<sup>(6)</sup>, dihydrogenistein (DHGE), and dihydrodaidzein (DHDE), and equol<sup>(28)</sup>, which may exert biological effects in addition to their parent compounds (i.e., daidzein and

genistein)<sup>(25)</sup>. In this regard, use of isoflavone metabolites in blood or urine as objective markers of isoflavone intake is an appealing approach<sup>(17; 32)</sup>.

In the current investigation we utilized data from a combined cohort based on two well-characterized cohorts of U.S. women, the Nurses' Health Study (NHS) and NHSII, to prospectively evaluate the association of urinary excretion of isoflavone metabolites with risk of T2D. We also examined the hypothesis that isoflavone excretion may especially be associated with lower T2D risk among postmenopausal women who do not receive replacement therapy and thus have a low exposure to exogenous estrogens.

## **METHODS**

### **Study population**

The NHS began in 1976, when 121,700 female registered nurses aged 30-55 y residing in 11 states were enrolled and completed a baseline questionnaire about their lifestyle and medical history. The NHSII was established in 1989 and consisted of 116,430 younger female registered nurses aged 25-42 y at baseline. These nurses also responded to a baseline questionnaire similar to the NHS. In both cohorts, questionnaires were collected at baseline and biennially thereafter, to update information on age, weight, smoking status, physical activity, medication use, menopausal status, postmenopausal hormone use, and disease status, including hypertension, hypercholesterolemia, cardiovascular disease (CVD), and cancer.

### **Urine sample collection**

A total of 18,743 NHS participants aged 53 to 80 years provided blood and urine samples from 2000 to 2002, and 29,611 NHSII participants aged 32 to 52 years provided blood and urine



samples from 1996 to 1999. For both cohorts the samples were returned to a central biorepository via overnight courier and were immediately processed upon arrival and aliquoted into cryotubes, which were stored in the vapor phase of liquid nitrogen freezers at  $\leq -130$  °C. Loss to follow-up was  $< 10\%$  among participants who provided blood and urine samples.

### **Prospective case-control study design**

We conducted prospective, nested case-control studies among participants who provided urine samples and were free of self-reported diabetes, CVD, and cancer at urine collection in NHS and NHSII separately. T2D cases diagnosed within the first year since urine sample collection were excluded from selection in order to reduce the potential for reverse causation bias. During follow-up from urine collection through 2008 (NHS)/2007 (NHSII), we prospectively identified and confirmed 1,111 T2D cases (NHS: 456; NHSII: 655) and randomly selected one control for each case.<sup>29</sup> The cases and controls were matched for age at sample collection, month of sample collection, fasting status ( $\geq 8$  h or not), first morning urine (yes or no), and race (white or other races) in both cohorts. In NHSII, we additionally matched for menopausal status (yes, no), luteal day of the menstrual cycle (date of next period minus date of sample collection) for premenopausal women, and hormone replacement therapy (yes or no) for postmenopausal women. The study protocol was approved by the institutional review board of the Brigham and Women's Hospital and the Human Subjects Committee Review Board of Harvard School of Public Health.

### **Ascertainment of T2D**

We sent a validated supplemental questionnaire to those who reported a physician diagnosis of T2D to confirm the incidence<sup>(16)</sup>. We used at least one of the following American Diabetes

Association 1998 criteria to confirm self-reported T2D diagnosis: (1) an elevated glucose concentration (fasting plasma glucose  $\geq 7.0$  mmol/l, random plasma glucose  $\geq 11.1$  mmol/l, or plasma glucose  $\geq 11.1$  mmol/l after an oral glucose load), and at least one symptom related to diabetes; no symptoms, but elevated glucose concentrations on two separate occasions; or treatment with insulin or oral hypoglycemic medication. Only confirmed T2D cases were included in the current study.

### **Assessment of diet**

Validated FFQs have been administered since 1984 in NHS or 1991 in NHSII<sup>38</sup>. Similar FFQs were subsequently sent to participants every two to four years to update diet. In these FFQs, we inquired about the consumption frequency of 118-166 food items in the past year and how often (from “never or less than once per month” to “6 or more times per day”) on average they consumed each food item of a standard portion size. Major soy foods, i.e., tofu and soy milk, have been simultaneously included on the FFQs since 1998 in the NHS and 1999 in the NHS II. An overall measure of diet quality was calculated by using the Alternate Healthy Eating Index (AHEI) score excluding the soy food items, i.e., tofu and soy milk<sup>(18)</sup>. In calculating AHEI score, we assigned individual score to each of the food group as a priority based on their beneficial effect to health, and summed up all of the scores.

### **Laboratory measurements**

In the current study, we used electrospray ionization liquid chromatography mass spectrometry to measure isoflavonoids in urine samples, which has been validated except for the use of an orbitrap mass spectrometer<sup>(6; 7)</sup>. Urinary creatinine levels were measured using a Roche-Cobas MiraPlus clinical chemistry autoanalyzer (Roche Diagnostics). The average intra-assay

coefficient of variation was 4.1% for daidzein, 7.6% for genistein, 8.2% for DHDE, 10.1% for DHGE, 8.1% for DMA, and 5.6% for creatinine. We calculated creatinine-adjusted concentrations (nmol/g creatinine) of isoflavonoids by dividing the isoflavonoid levels (nmol/L) by creatinine levels (g/L). In a pilot study that evaluated within-person stability of the isoflavonoids, intra-class correlation coefficients (ICCs), of two urine samples from 58 NHSII participants collected 1-2 years apart were 0.05 for daidzein, 0.14 for genistein, 0.16 for DHDE, 0.14 for DHGE, and 0.20 for DMA.

### **Statistical methods**

We calculated Spearman correlation coefficients between urine excretion of isoflavone metabolites with soy foods, i.e., tofu and soy milk, estimated from FFQ. We adjusted for total energy intake (kcal/day), BMI ( $\text{kg/m}^2$ ), physical activity (MET-hr/week), age (years), smoking (never, past, current), and first morning urine (yes or no). This analysis was conducted among controls to facilitate comparison between cohorts.

We categorized the isoflavones biomarkers into tertiles. We used conditional logistic regression stratified by matching factors to model the association between isoflavone metabolites and risk of T2D in the main analysis <sup>(29)</sup>. We additionally adjusted for hypertension at baseline (yes or no), hypercholesterolemia at baseline (yes or no), body mass index ( $\text{kg/m}^2$ ), smoking (nonsmoker, past smoker, current smoker), AHEI score, physical activity (METs-hr/week), total energy intake (kcal/day), menopausal status (premenopausal or postmenopausal) (NHS only), and hormone replacement therapy (yes or no) (NHS only). P values for linear trend were calculated by examining an ordinal score based on the median value in each tertile of isoflavones biomarker levels in the multivariate models.

Given that menopausal status and postmenopausal hormone therapy were not matching factors in NHS, we conducted stratified analyses by menopausal status and postmenopausal hormone therapy (yes vs. no) using unconditional logistic regression to maximize statistical power. P values for interactions were evaluated using likelihood ratio test, comparing the multivariate model with and without interaction terms of dichotomized isoflavones and potential effect modifiers in conditional logistic regression. Joint associations of the urinary biomarkers and potential effect modifiers were estimated using conditional logistic regression. All P values were two-sided. Data were analyzed with the Statistical Analysis Systems software package, version 9.3 (SAS Institute, Inc., Cary, North Carolina).

## RESULTS

Baseline characteristics of study participants are shown in **Table 1**. Compared with controls, T2D cases consumed a less healthful diet, engaged in less physical activity, had a higher BMI, and were more likely to have a history of hypertension and hypercholesterolemia. The baseline characteristics according to urine isoflavone excretion are shown in **Supplemental Table 1**.

Higher urinary isoflavone excretion was correlated with a healthier dietary pattern, and higher levels of physical activity.

Moderate to strong correlations among urinary isoflavone metabolites were observed (correlation coefficient: 0.26 - 0.79), with the strongest correlation observed between daidzein and genistein (**Supplemental table 2**). Weak yet significant correlations were found between soy foods estimated from FFQ and urinary isoflavone excretion among controls.

Urinary excretion of daidzein and genistein, which are the main components of isoflavones, was associated with a lower risk of T2D (**Table 2**). For daidzein, compared with the lowest group, the odds ratios (ORs) of T2D were 0.78 (0.60, 1.02) in the second tertile and 0.71 (0.55, 0.93) in the highest tertile (P for trend = 0.03); for genistein, compared with the lowest group, the ORs of T2D were 0.70 (0.53, 0.93) in the second tertile and 0.74 (0.56, 0.97) in the highest tertile (P for trend = 0.15). DMA, DHDE, and DHGE, which are the metabolites of daidzein and genistein, were non-significantly associated with lower risk of T2D. Comparing the extreme tertiles, the odds ratios of T2D was 0.92 (95% CI: 0.70, 1.21) for DMA, 0.80 (95% CI: 0.60, 1.06) for DHDE, and 0.82 (95% CI: 0.62, 1.08) for DHGE, respectively.

Stratified analyses by menopausal status and hormone use were conducted. Significant interaction by postmenopausal hormone use was found for daidzein (P for interaction = 0.001) (**Table 3**). Comparing extreme tertiles of the urinary daidzein, the ORs of T2D was 0.81 (95% CI: 0.54, 1.21; P for trend = 0.41) for premenopausal women, 0.76 (95% CI: 0.52, 1.11; P for trend = 0.19) for postmenopausal women with hormone use, and 0.58 (95% CI: 0.34, 0.97; P for trend = 0.06) for postmenopausal women without hormone use. Consistently, in joint association analysis the inverse association between urinary daidzein and risk of T2D appeared to be stronger in the postmenopausal women without hormone use (**Figure 1**). No significant interactions between other metabolites and risk of T2D by menopausal status and hormone use were found. Further stratified analyses were conducted by age, BMI, and aHEI, and the inverse association appeared to be more apparent among women with BMI less than 30 (P for interaction = 0.03) (**Supplemental table 3**).

## DISCUSSION

In the combined cohorts of U.S. women, the main components of urinary isoflavones, daidzein and genistein, were associated with a lower risk of T2D. Further analyses suggested that inverse association between daidzein and risk of T2D appeared to be stronger among postmenopausal women who did not take hormone therapy at sample collection. These associations were independent of established diabetes risk factors, such as BMI, physical activity, and overall diet quality.

Isoflavones are able to bind to the **estrogen** receptors (ER), especially ER- $\beta$ , with  $10^3$ - $10^4$  less potency than estradiol <sup>(25)</sup>. These compounds can exert either estrogenic or anti-estrogenic action depending on the level of estradiol in the circulation. When endogenous estrogen levels are low, isoflavones primarily exert estrogen-like effects <sup>(19)</sup>. Our observation that a stronger inverse association between isoflavones, especially daidzein, and risk of T2D was noted among postmenopausal women without current hormone use is in line with the notion that isoflavones exert estrogen-like effects on blood glucose when circulating estrogen levels are low. In addition, this inverse association was also consistent with results from short-term clinical trials.

Supplementation of isoflavones did not significantly lower fasting glucose or insulin levels in a comprehensive meta-analysis of randomized clinical trials <sup>(12)</sup>. Although the non-significant results of those trials might be due to short term duration and might not represent the anti-diabetic potential of long-term isoflavone intake, however, in another meta-analysis that focused on perimenopausal and postmenopausal non-Asian women who did not take hormone replacement therapy, isoflavone supplementation significantly lowered fasting insulin and HOMA-IR levels <sup>(30)</sup>. Meanwhile, a stronger inverse association between urinary daidzein and risk of T2D was observed among non-obese participants. The reason might be that blood circulating estrogen might be lower in non-obese women comparing to obese women, as adipose

tissue is the main source of circulating estrogens for postmenopausal women without hormone use<sup>(13)</sup>.

Besides binding to ER, isoflavones may also bind to and activate nuclear receptors which regulate lipid and glucose metabolism, including liver X binding receptor (LXR), sterol regulated element binding protein (SREBP), peroxisome-proliferator activated receptors  $\alpha$  (PPAR $\alpha$ ), and PPAR $\gamma$ <sup>(20; 11; 21)</sup>. Isoflavones may also increase the phosphorylation of AMP-activated protein kinase (AMPK) and acetyl-CoA carboxylase to improve glucose uptake and fatty acid oxidation<sup>4</sup>. Animal studies have also shown that isoflavones may improve hyperglycemia, glucose tolerance, and circulating insulin concentrations<sup>(2)</sup>.

Several cohort studies have been conducted to examine the association between intakes of soy foods and isoflavones estimated using FFQs and risk of T2D<sup>(24 37; 26; 22; 40)</sup>. For example, in a Japanese population, higher intakes of isoflavones were not associated with diabetes risk in the total study population, but an inverse association was found among overweight women<sup>(26)</sup>. In the Singapore Chinese Health Study, intakes of total isoflavones were not associated with T2D risk, probably because many soy foods were sweetened in Singapore<sup>(24)</sup>. In the EPIC-InterAct Study, total isoflavone intake was not associated with T2D risk among men and women in eight European countries<sup>(40)</sup>. In addition, two other studies assessed soy food consumption with risk of T2D. An inverse association was found in a Chinese cohort<sup>(37)</sup>, while a positive association found in the Multiethnic Cohort in Hawaii<sup>(22)</sup>. Differences in processing and cooking methods of soy food might be an explanation of these inconsistent associations. Moreover, measurement errors of FFQ assessments may be of a particular concern, especially among Western populations who consume much less soy foods than Asians. To our knowledge, no previous prospective studies have examined isoflavone biomarkers in relation to T2D risk. In cross-sectional studies

on the association between urinary isoflavones and blood glucose concentration, no associations were found<sup>(27; 34)</sup>. The lack of association might be due to the reverse causation bias that the diabetes participants might tend to increase consumption of plant-based diet containing soy foods.

Our study had several strengths, including large sample size, long follow-up duration, and use of urinary biomarkers to represent the internal dose of isoflavones and account for the inter-individual variations in bioavailability. Our study also has several limitations. First, the isoflavones were measured in spot urine samples, and the frequency of soy food consumption in our population was very low. Thus, we obtained low ICCs, which suggested that urinary biomarkers might not well represent long-term excretion of isoflavones.. However, in general, non-differential misclassification of the exposure biases the results towards null. Moreover, only moderate correlation was found between soy foods assessed by FFQ and urinary isoflavones, showing that urinary isoflavones might not well represent daily soy food consumption. The reason might be due to the measurement errors in both spot urine and FFQ. Second, although we adjusted for an array of established and potential risk factors of T2D, we could not exclude the possibility that unmeasured confounding or residual confounding, such as a healthy lifestyle, might still partially explain the association between isoflavones and risk of T2D. Last, we cannot exclude the possibility that chance may play a role in the inverse association only found between daidzein and risk of T2D.

In conclusion, we observed inverse associations between urinary excretion of daidzein and genistein and risk of T2D in U.S. women. In addition, the inverse association for daidzein appeared to be stronger among postmenopausal women who did not use hormone replacement therapy. Although these findings are in line with evidence from animal experiments and clinical trials demonstrating benefits of isoflavone intake on insulin resistance, further studies are



warranted to replicate the current findings, preferably using multiple-day 24-hr urine samples to achieve more stable estimates of isoflavone marker excretion.

### **Contributors**

RMvD and FBH obtained funding from the National Institutes of Health. QS, AAF, FBH, and RMvD were involved in data collection of dietary flavonoids or urinary metabolites. AAF measured urinary metabolites using liquid chromatography mass spectrometry. QS and RMvD conducted pilot studies for the current investigation. QS, FBH, BAR, and RMvD provided statistical expertise. MD analyzed the data and wrote the first draft of the manuscript. QS, FBH, AAF, SST, BAR, and RMvD contributed to the interpretation of the results and critical revision of the manuscript for important intellectual content and approved the final version of the manuscript. QS is the guarantor of this investigation.

### **Funding**

This study was funded by research grants CA186107, CA49449, CA176726, CA67262, DK58845, DK58785, DK082486, CA50385, CA87969, and P30 CA71789 from the National Institutes of Health. Dr. Sun was supported by a career development award R00HL098459 from the National Heart, Lung, and Blood Institute. The funding sources had no role in study design; in the collection, analysis, and interpretation of data; in the writing of the report; or in the decision to submit the article for publication. The authors are not affiliated with the funding institutions.

### **Competing Interests**

None of the authors had any financial or personal conflict of interest to disclose.

## Acknowledgement

We acknowledge Dr. Mary Kay Townsend and Dr. Kimberly A. Bertrand's work on conducting pilot studies for the current investigation.

## Reference

1. (2013) International Diabetes Federation. IDF Diabetes Atlas, 6th edn. *Brussels, Belgium: International Diabetes Federation*.
2. Babu PV, Liu D & Gilbert ER (2013) Recent advances in understanding the anti-diabetic actions of dietary flavonoids. *J Nutr Biochem* **24**, 1777-1789.
3. Bhagwat S HD, Holden JM, et al. (2008) USDA Database for the Isoflavone Content of Selected Foods Release 2.0.
4. Cederroth CR, Vinciguerra M, Gjinojci A *et al.* (2008) Dietary phytoestrogens activate AMP-activated protein kinase with improvement in lipid and glucose metabolism. *Diabetes* **57**, 1176-1185.
5. Cho GJ, Lee JH, Park HT *et al.* (2008) Postmenopausal status according to years since menopause as an independent risk factor for the metabolic syndrome. *Menopause* **15**, 524-529.
6. Franke AA, Custer LJ, Wilkens LR *et al.* (2002) Liquid chromatographic-photodiode array mass spectrometric analysis of dietary phytoestrogens from human urine and blood. *J Chromatogr B Analyt Technol Biomed Life Sci* **777**, 45-59.
7. Franke AA, Halm BM, Kakazu K *et al.* (2009) Phytoestrogenic isoflavonoids in epidemiologic and clinical research. *Drug Test Anal* **1**, 14-21.
8. Fung TT, Schulze M, Manson JE *et al.* (2004) Dietary patterns, meat intake, and the risk of type 2 diabetes in women. *Arch Intern Med* **164**, 2235-2240.
9. Heianza Y, Arase Y, Kodama S *et al.* (2013) Effect of postmenopausal status and age at menopause on type 2 diabetes and prediabetes in Japanese individuals: Toranomon Hospital Health Management Center Study 17 (TOPICS 17). *Diabetes Care* **36**, 4007-4014.
10. Kanaya AM, Herrington D, Vittinghoff E *et al.* (2003) Glycemic effects of postmenopausal hormone therapy: the Heart and Estrogen/progestin Replacement Study. A randomized, double-blind, placebo-controlled trial. *Ann Intern Med* **138**, 1-9.
11. Kim S, Sohn I, Ahn JI *et al.* (2004) Hepatic gene expression profiles in a long-term high-fat diet-induced obesity mouse model. *Gene* **340**, 99-109.
12. Liu ZM, Chen YM & Ho SC (2011) Effects of soy intake on glycemic control: a meta-analysis of randomized controlled trials. *Am J Clin Nutr* **93**, 1092-1101.
13. Maccio A & Madeddu C (2011) Obesity, inflammation, and postmenopausal breast cancer: therapeutic implications. *ScientificWorldJournal* **11**, 2020-2036.
14. Manson JE, Chlebowski RT, Stefanick ML *et al.* (2013) Menopausal hormone therapy and health outcomes during the intervention and extended poststopping phases of the Women's Health Initiative randomized trials. *JAMA* **310**, 1353-1368.

15. Manson JE, Rimm EB, Colditz GA *et al.* (1992) A prospective study of postmenopausal estrogen therapy and subsequent incidence of non-insulin-dependent diabetes mellitus. *Ann Epidemiol* **2**, 665-673.
16. Manson JE, Rimm EB, Stampfer MJ *et al.* (1991) Physical activity and incidence of non-insulin-dependent diabetes mellitus in women. *Lancet* **338**, 774-778.
17. Maskarinec G, Singh S, Meng L *et al.* (1998) Dietary soy intake and urinary isoflavone excretion among women from a multiethnic population. *Cancer Epidemiol Biomarkers Prev* **7**, 613-619.
18. McCullough ML & Willett WC (2006) Evaluating adherence to recommended diets in adults: the Alternate Healthy Eating Index. *Public Health Nutr* **9**, 152-157.
19. Messina MJ & Wood CE (2008) Soy isoflavones, estrogen therapy, and breast cancer risk: analysis and commentary. *Nutr J* **7**, 17.
20. Mezei O, Banz WJ, Steger RW *et al.* (2003) Soy isoflavones exert antidiabetic and hypolipidemic effects through the PPAR pathways in obese Zucker rats and murine RAW 264.7 cells. *J Nutr* **133**, 1238-1243.
21. Mezei O, Li Y, Mullen E *et al.* (2006) Dietary isoflavone supplementation modulates lipid metabolism via PPARalpha-dependent and -independent mechanisms. *Physiol Genomics* **26**, 8-14.
22. Morimoto Y, Steinbrecher A, Kolonel LN *et al.* (2011) Soy consumption is not protective against diabetes in Hawaii: the Multiethnic Cohort. *Eur J Clin Nutr* **65**, 279-282.
23. Mozaffarian D, Hao T, Rimm EB *et al.* (2011) Changes in diet and lifestyle and long-term weight gain in women and men. *N Engl J Med* **364**, 2392-2404.
24. Mueller NT, Odegaard AO, Gross MD *et al.* (2012) Soy intake and risk of type 2 diabetes in Chinese Singaporeans [corrected]. *Eur J Nutr* **51**, 1033-1040.
25. Muthyala RS, Ju YH, Sheng S *et al.* (2004) Equol, a natural estrogenic metabolite from soy isoflavones: convenient preparation and resolution of R- and S-equols and their differing binding and biological activity through estrogen receptors alpha and beta. *Bioorg Med Chem* **12**, 1559-1567.
26. Nanri A, Mizoue T, Takahashi Y *et al.* (2010) Soy product and isoflavone intakes are associated with a lower risk of type 2 diabetes in overweight Japanese women. *J Nutr* **140**, 580-586.
27. Penalvo JL & Lopez-Romero P (2012) Urinary enterolignan concentrations are positively associated with serum HDL cholesterol and negatively associated with serum triglycerides in U.S. adults. *J Nutr* **142**, 751-756.
28. Perez-Jimenez J, Hubert J, Hooper L *et al.* (2010) Urinary metabolites as biomarkers of polyphenol intake in humans: a systematic review. *Am J Clin Nutr* **92**, 801-809.
29. Prentice RL & Breslow NE (1978) Retrospective studies and failure time models. *Biometrika* **65**, 153-158.
30. Ricci E, Cipriani S, Chiaffarino F *et al.* (2010) Effects of soy isoflavones and genistein on glucose metabolism in perimenopausal and postmenopausal non-Asian women: A meta-analysis of randomized controlled trials. *Menopause* **17**, 1080-1086.
31. Rossi R, Origliani G & Modena MG (2004) Transdermal 17-beta-estradiol and risk of developing type 2 diabetes in a population of healthy, nonobese postmenopausal women. *Diabetes care* **27**, 645-649.
32. Seow A, Shi CY, Franke AA *et al.* (1998) Isoflavonoid levels in spot urine are associated with frequency of dietary soy intake in a population-based sample of middle-aged and older Chinese in Singapore. *Cancer Epidemiol Biomarkers Prev* **7**, 135-140.
33. Spencer JP, Abd El Mohsen MM, Minihaane AM *et al.* (2008) Biomarkers of the intake of dietary polyphenols: strengths, limitations and application in nutrition research. *Br J Nutr* **99**, 12-22.
34. Struja T, Richard A, Linseisen J *et al.* (2014) The association between urinary phytoestrogen excretion and components of the metabolic syndrome in NHANES. *Eur J Nutr* **53**, 1371-1381.
35. Szmilowicz ED, Stuenkel CA & Seely EW (2009) Influence of menopause on diabetes and diabetes risk. *Nat Rev Endocrinol* **5**, 553-558.

36. van Dam RM, Rimm EB, Willett WC *et al.* (2002) Dietary patterns and risk for type 2 diabetes mellitus in U.S. men. *Ann Intern Med* **136**, 201-209.
37. Villegas R, Gao YT, Yang G *et al.* (2008) Legume and soy food intake and the incidence of type 2 diabetes in the Shanghai Women's Health Study. *Am J Clin Nutr* **87**, 162-167.
38. Willett WC (1998) *Nutritional epidemiology*. 2 ed: New York: Oxford University Press.
39. Yang B, Chen Y, Xu T *et al.* (2011) Systematic review and meta-analysis of soy products consumption in patients with type 2 diabetes mellitus. *Asia Pac J Clin Nutr* **20**, 593-602.
40. Zamora-Ros R, Forouhi NG, Sharp SJ *et al.* (2013) The association between dietary flavonoid and lignan intakes and incident type 2 diabetes in European populations: the EPIC-InterAct study. *Diabetes Care* **36**, 3961-3970.
41. Zamora-Ros R, Touillaud M, Rothwell JA *et al.* (2014) Measuring exposure to the polyphenol metabolome in observational epidemiologic studies: current tools and applications and their limits. *Am J Clin Nutr* **100**, 11-26.

Table 1. The age-adjusted baseline characteristics according to diabetes cases and controls in the combined cohort

Characteristics*	Case	Control
	(N=1111)	(N=1111)
Age at urine collection (year)	53.4	53.4
Body mass index (kg/m <sup>2</sup> )	32.0	25.9
Physical Activity (MET-hr/week)	16.4	19.5
Current smoker (%)	11	7
Hypertension (%)	38	19
Hypercholesterolemia (%)	55	35
Family history of diabetes (%)	37	21
First morning urine (%)‡	88	88
Postmenopausal (%)‡	83	83
Postmenopausal hormone use (% of postmenopausal women) ‡	43	43
White (%)‡	96	97
<i>Urinary metabolites (quartile range)</i> (nmol/g creatinine) #		
Daidzein	311 (112, 956)	343 (123, 1013)
Genistein	122 (48, 424)	128 (48, 419)
Desmethylangolensin	29 (9, 116)	36 (12, 135)
Dihydrodaidzein	34 (6, 217)	42 (9, 288)
Dihydrogenistein	25 (10, 72)	27 (12, 78)
<i>Diet</i>		
Total energy (kcal/day)	1865	1775
Alcohol (g/d)	0.3	0.4
Coffee (cup/day)	1.0	1.2
Soft drinks (serving/day)	1.2	0.9
Tofu consumers (%)	10	12

Soy milk consumers (%)	4	7
Fruits (serving/day)	2.0	2.0
Vegetables (serving/day)	3.5	3.3
Red meat (serving/day)	1.5	1.3
Fish (serving/day)	0.2	0.2
Hot dog (serving/day)	0.1	0.1
Alternate Healthy Eating Index score	49	52

\*All of the variables were age-adjusted, except the urinary metabolites of isoflavones. Values of continuous variables were median. Percentages were based on non-missing data.

‡Matching factors; menopausal status and hormone replacement therapy were matching factors for NHSII only.

#Values were median (interquartile range).

Table 2. Odds ratio (95% CI) of type 2 diabetes by tertiles of urinary isoflavones (nmol/g creatinine) in the combined cohort

	Tertiles of urinary markers			P for trend
	1 (lowest)	2	3 (highest)	
<b>Daidzein</b>				
Median	77	326	1,529	
(Range)	(0, 171)	(171, 665)	(665, 196,860)	
Case/control	386/354	369/372	356/385	
Model 1	1.00	0.88 (0.72, 1.07)	0.85 (0.70, 1.04)	0.18
Model 2	1.00	0.78 (0.60, 1.02)	0.71 (0.55, 0.93)	0.03
<b>Genistein</b>				
Median	33	125	729	
(Range)	(0, 65)	(65, 269)	(270, 375,579)	
Case/control	376/364	370/371	365/376	
Model 1	1.00	0.94 (0.76, 1.15)	0.92 (0.74, 1.13)	0.50
Model 2	1.00	0.70 (0.53, 0.93)	0.74 (0.56, 0.97)	0.15
<b>Desmethylangolensin (DMA)</b>				
Median	6	33	225	
(Range)	(0, 16)	(16, 78)	(78, 160,620)	
Case/control	393/347	365/376	353/388	
Model 1	1.00	0.82 (0.68, 1.01)	0.77 (0.63, 0.94)	0.039
Model 2	1.00	1.00 (0.77, 1.31)	0.92 (0.70, 1.21)	0.74
<b>Dihydrodaidzein (DHDE)</b>				
Median	2	38	568	
(Range)	(0, 14)	(14, 126)	(126, 61,237)	
Case/control	392/348	366/375	353/388	

Model 1	1.00	0.90 (0.73, 1.10)	0.78 (0.63, 0.97)	0.043
Model 2	1.00	0.94 (0.72, 1.23)	0.80 (0.60, 1.06)	0.16
<b>Dihydrogenistein (DHGE)</b>				
Median	7	26	150	
(Range)	(0, 15)	(15, 49)	(49, 64,372)	
Case/control	389/351	357/384	365/376	
Model 1	1.00	0.84 (0.68, 1.02)	0.89 (0.72, 1.09)	0.55
Model 2	1.00	0.73 (0.56, 0.96)	0.82 (0.62, 1.08)	0.45

Model 1: Conditional logistic model stratified by matching factors including age at urine sample collection, month of sample collection, first morning urine (yes or no), and race (white or other races) in both cohorts. Menopausal status (premenopausal or postmenopausal), and hormone replacement therapy (yes or no) were additionally matched for in NHS II.

Model 2: Conditional logistic model additionally adjusted for family history of diabetes (yes or no), hypertension at baseline (yes or no), hypercholesterolemia at baseline (yes or no), body mass index ( $\text{kg/m}^2$ ), smoking (nonsmoker, past smoker, current smoker), alternative healthy eating index, physical activity (MET-hr/week), total energy intake (kcal/day), menopausal status (premenopausal or postmenopausal), and hormone replacement therapy (yes or no).



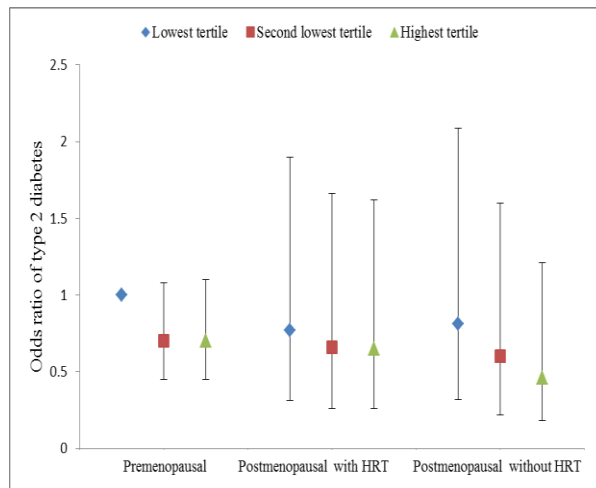
Table 3. Stratified analysis of the association between urine isoflavones biomarkers and risk of type 2 diabetes by menopausal status and postmenopausal hormone use in the combined cohort

	Number of participants	T1	T2	T3	P for interaction*
<b>Daidzein</b>					
Premenopausal	975	1.00	0.85 (0.58, 1.24)	0.81 (0.54, 1.21)	0.001
Postmenopausal with HRT	821	1.00	0.86 (0.59, 1.25)	0.76 (0.52, 1.11)	
Postmenopausal without HRT	426	1.00	0.70 (0.41, 1.19)	0.58 (0.34, 0.97)	
<b>Genistein</b>					
Premenopausal	975	1.00	0.96 (0.66, 1.41)	1.03 (0.70, 1.52)	0.36
Postmenopausal with HRT	821	1.00	0.73 (0.50, 1.07)	0.75 (0.52, 1.10)	
Postmenopausal without HRT	426	1.00	0.58 (0.34, 1.00)	0.51 (0.30, 0.86)	
<b>DMA</b>					
Premenopausal	975	1.00	1.19 (0.80, 1.77)	1.27 (0.86, 1.87)	0.21
Postmenopausal with HRT	821	1.00	1.06 (0.73, 1.53)	0.74 (0.51, 1.09)	
Postmenopausal without HRT	426	1.00	0.79 (0.47, 1.34)	0.77 (0.45, 1.33)	
<b>DHDE</b>					
Premenopausal	975	1.00	1.22 (0.82, 1.80)	1.22 (0.82, 1.82)	0.82
Postmenopausal with HRT	821	1.00	0.99 (0.68, 1.43)	0.65 (0.44, 0.96)	
Postmenopausal without HRT	426	1.00	0.75 (0.44, 1.28)	0.63 (0.37, 1.05)	
<b>DHGE</b>					
Premenopausal	975	1.00	0.70 (0.48, 1.03)	1.03 (0.70, 1.51)	0.90
Postmenopausal with HRT	821	1.00	0.84 (0.57, 1.23)	0.66 (0.45, 0.96)	
Postmenopausal without HRT	426	1.00	0.80 (0.47, 1.36)	0.71 (0.42, 1.22)	

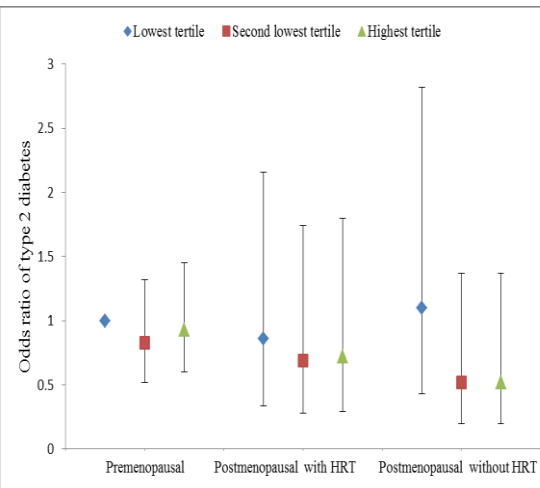
HRT: hormone replacement therapy

Given that menopausal status and HRT were not matching factors in NHS, unconditional logistic models were used for the stratified analysis, and multivariate models adjusted for age at urine sample collection, month of sample collection, fasting status ( $\geq 8$  h or not), first morning urine (yes or no), race (white or other races), family history of diabetes (yes or no), hypertension at baseline (yes or no), hypercholesterolemia at baseline (yes or no), body mass index (continuous), smoking (nonsmoker, past smoker, current smoker), alternative healthy eating index (continuous), physical activity (continuous), total energy intake (continuous).

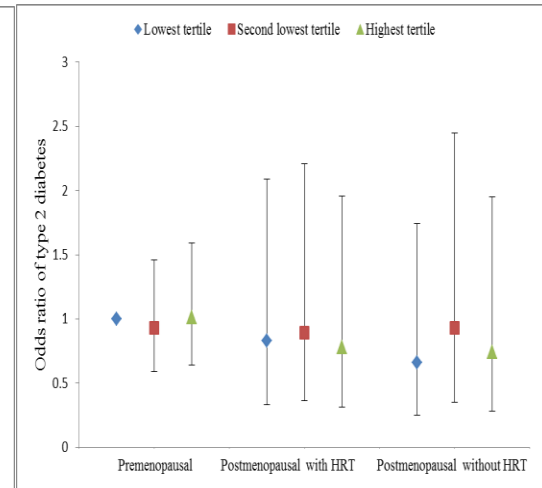
\* Likelihood ratio test comparing the conditional logistic models with and without interaction term was used; the urine isoflavones were dichotomized, and the premenopausal and postmenopausal with HRT were combined into one group.



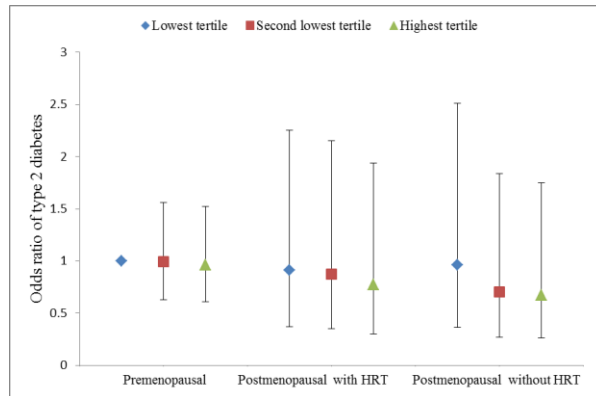
1a. Daidzein



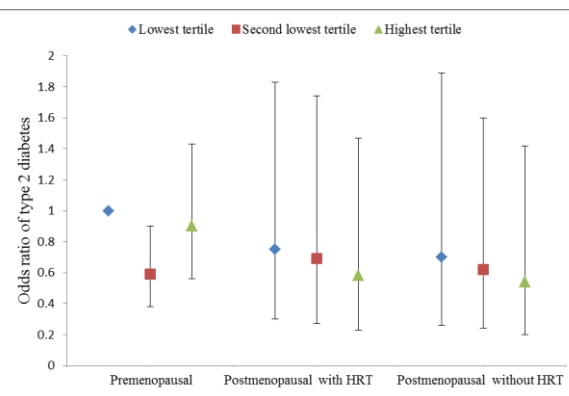
1b. Genestein



1c. DMA



1d. DHDE



1e. DHGE

Figure 1. The joint association of urinary isoflavones biomarkers and postmenopausal status and hormone use with risk of type 2 diabetes.

Conditional logistic models adjusted for hypertension at baseline (yes or no), hypercholesterolemia at baseline (yes or no), body mass index ( $\text{kg}/\text{m}^2$ ), smoking (nonsmoker, past smoker, current smoker), alternative healthy eating index, physical activity (MET-hr/week), and total energy intake (kcal/day).

Supplemental Table 1. The age-adjusted baseline characteristics according to urinary total isoflavones in the controls of the combined cohort

Characteristics *	T1	T2	T3
Number	361	371	379
Urinary metabolites (nmol/g creatinine)			
Daidzein	102	424	4766
Genistein	49	184	5188
DHDE	17	101	2686
DHGE	20	46	987
DMA	22	72	1589
Diet			
Total energy (kcal/day)	1743	1817	1755
Alcohol (g/d)	0.4	0.3	0.3
Coffee (cup/day)	1.2	1.2	1.2
Soft drinks (serving/day)	1.0	1.0	0.8
Tofu consumers (%)	10	13	17
Soy milk consumers (%)	6	5	11
Fruits (serving/day)	2.0	2.0	2.0
Vegetables (serving/day)	3.2	3.3	3.4
Red meat (serving/day)	1.3	1.4	1.2
Fish (serving/day)	0.2	0.2	0.2
Hotdog (serving/day)	0.1	0.1	0.1
Alternate Healthy Eating Index score	52	50	52
Age at urine sample	53	54	54

---

collection (year)			
Hypertension (%)	16	17	21
Hypercholesterolemia (%)	33	29	36
Family history of diabetes (%)	22	22	18
First morning urine (%)	85	88	90
Postmenopausal (%)	81	82	80
Postmenopausal hormone use among postmenopausal women (%)	41	41	40
Current smoker (%)	7	9	7
White (%)	98	96	97
Body mass index (kg/m <sup>2</sup> )	26	26	26
Physical Activity (MET-hr/week)	22	18	19

---

Supplemental table 2. Spearman correlation coefficient<sup>†</sup> between urine isoflavone biomarkers and soy food products assessed by FFQ among controls in the combined cohort

	Daidzein	Genistein	DMA	DHDE	DHGE
Daidzein	1.00				
Genistein	0.79*	1.00			
DMA	0.60*	0.39*	1.00		
DHDE	0.61*	0.56*	0.60*	1.00	
DHGE	0.43*	0.55*	0.26*	0.58*	1.00
Soy milk (FFQ)	0.10*	0.14*	0.11*	0.10*	0.11*
Tofu (FFQ)	0.15*	0.16*	0.11*	0.13*	0.11*
Total soy food	0.15*	0.18*	0.12*	0.14*	0.13*
Isoflavones calculated from foods	0.09*	0.11*	0.12*	0.11*	0.12*

\*P<0.05

<sup>†</sup>Spearman correlation coefficient adjusted for total energy intake (kcal/day), BMI (kg/m<sup>2</sup>), physical activity (MET-hr/week), age (years), and smoking (never, past, current).

Soy foods were the average amount of consumption estimated from FFQs in 1998 and 2002 for NHS, and in 1995 and 1999 for NHS2.

Supplemental Table 3. Stratified analysis of the association between urine isoflavones biomarkers and risk of type 2 diabetes by age, BMI, and AHEI

	N	T1	T2	T3	P for interaction*
<b>Urinary daidzein</b>					
<i>Age</i>					
Age > 60 y	691	1.00	0.68 (0.46, 1.02)	0.65 (0.44, 0.97)	0.07
Age ≤ 60 y	1531	1.00	0.92 (0.69, 1.24)	0.80 (0.59, 1.08)	
<i>BMI</i>					
BMI < 30 kg/m <sup>2</sup>	1406	1.00	0.84 (0.64, 1.11)	0.63 (0.47, 0.84)	0.03
BMI ≥ 30 kg/m <sup>2</sup>	816	1.00	0.87 (0.58, 1.31)	1.05 (0.69, 1.58)	
<i>aHEI</i>					
aHEI < 52	1233	1.00	0.76 (0.55, 1.05)	0.58 (0.42, 0.81)	0.48
aHEI ≥ 52	989	1.00	0.93 (0.65, 1.31)	0.98 (0.70, 1.39)	
<b>Urinary genistein</b>					
<i>Age</i>					
Age > 60 y	691	1.00	0.70 (0.46, 1.04)	0.75 (0.50, 1.11)	0.80
Age ≤ 60 y	1531	1.00	0.82 (0.61, 1.10)	0.82 (0.61, 1.11)	
<i>BMI</i>					
BMI < 30 kg/m <sup>2</sup>	1406	1.00	0.74 (0.56, 0.98)	0.71 (0.54, 0.95)	0.16
BMI ≥ 30 kg/m <sup>2</sup>	816	1.00	1.00 (0.67, 1.51)	0.96 (0.64, 1.45)	
<i>aHEI</i>					
aHEI < 52	1233	1.00	0.70 (0.51, 0.96)	0.63 (0.45, 0.87)	0.35
aHEI ≥ 52	989	1.00	0.89 (0.62, 1.28)	1.04 (0.74, 1.47)	

AHEI: alternative healthy eating index. AHEI ranged from 0 to 100, with a higher score indicating a healthier diet.

HRT: hormone replacement therapy

Unconditional logistic models were used for the stratified analysis, and multivariate models adjusted for age at urine sample collection, month of sample collection, fasting status (≥ 8 h or not), first morning urine (yes or no), race (white or other races), family history of



diabetes (yes or no), hypertension at baseline (yes or no), hypercholesterolemia at baseline (yes or no), body mass index (continuous), smoking (nonsmoker, past smoker, current smoker), alternative healthy eating index (continuous), physical activity (continuous), total energy intake (continuous).

\* Likelihood ratio test comparing the conditional logistic models with and without interaction term was used; the urine isoflavones were dichotomized.

## II CONCLUSION

---

Preventing obesity related chronic disease through promotion of a healthy diet is a public health priority globally. The work presented here shows that higher consumption of coffee and soy food may be beneficial to several health outcomes, especially type 2 diabetes and CVD.

In Chapter 1, I found a nonlinear relationship of coffee consumption with CVD risk: moderate coffee consumption was associated with lower risk of CVD, with the lowest CVD risk at 3 to 5 cups per day, and heavy coffee consumption was not associated with risk of CVD. In Chapter 2, similar to the findings in Chapter 1, I found a non-linear association of coffee consumption with total mortality in the whole population. When restricting to never smokers, coffee consumption was associated with lower risk of total mortality, and mortality due to CVD, neurological diseases, and suicide. The study provides strong evidence that the non-linear association of coffee with total mortality might be due to the confounding of smoking.

In Chapters 3 and 4, I found that consumption of isoflavones was associated with lower risk of type 2 diabetes in three large cohorts. Consistently, I further showed that higher concentrations of urinary isoflavones were associated with lower risk of type 2 diabetes.

Taken together, the findings of the four projects suggest that coffee and soy food can be incorporated into a healthy dietary pattern. However, randomized clinical trials are warranted to further examine the causal effects of coffee and soy food consumption on various health outcomes and explore biological mechanisms.